



157

Attorney Docket No. 522.1921D

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Patent Application of:

Mayumi TOMIKAWA et al.

Application No.: 09/910,054

Group Art Unit: 1631

Filed: July 23, 2001

Examiner: Moran, Marjorie A.

For: METHOD AND APPARATUS FOR EXTRACTING AND EVALUATING MUTUALLY  
SIMILAR PORTIONS IN ONE-DIMENSIONAL SEQUENCES IN MOLECULES AND/OR  
THREE-DIMENSIONAL STRUCTURES OF MOLECULES

**SUBMISSION OF CERTIFIED COPY OF PRIOR FOREIGN  
APPLICATION IN ACCORDANCE  
WITH THE REQUIREMENTS OF 37 C.F.R. § 1.55**

Commissioner for Patents  
PO Box 1450  
Alexandria, VA 22313-1450

Sir:

In accordance with the provisions of 37 C.F.R. § 1.55, the applicant(s) submit(s) herewith  
a certified copy of the following foreign application:

Japanese Patent Application No(s). 1992-331703

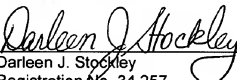
Filed: December 11, 1992

It is respectfully requested that the applicant(s) be given the benefit of the foreign filing  
date(s) as evidenced by the certified papers attached hereto, in accordance with the  
requirements of 35 U.S.C. § 119.

Respectfully submitted,  
STAAS & HALSEY LLP

Date: July 30, 2007

By:

  
Darleen J. Stockley  
Registration No. 34,257

1201 New York Ave, N.W., 7th Floor  
Washington, D.C. 20005  
Telephone: (202) 434-1500  
Facsimile: (202) 434-1501

# 日 本 国 特 許 庁

JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日  
Date of Application: 1 9 9 2 年 1 2 月 1 1 日

出 願 番 号  
Application Number: 平成 4 年特許願第 3 3 1 7 0 3 号

パリ条約による外国への出願  
用いる優先権の主張の基礎  
となる出願の国コードと出願  
号  
e country code and number  
your priority application,  
be used for filing abroad  
under the Paris Convention, is

J P 1 9 9 2 - 3 3 1 7 0 3

願 人  
Applicant(s): 富士通株式会社

2 0 0 7 年 7 月 6 日

特許庁長官  
Commissioner,  
Japan Patent Office

中 嶋



【書類名】 特許願

【整理番号】 9213453

【提出日】 平成 4年12月11日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 15/20

【発明の名称】 立体構造処理装置

【請求項の数】 19

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中 1 0 1 5 番地 富士通株式会社内

【氏名】 相川 聖一

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中 1 0 1 5 番地 富士通株式会社内

【氏名】 松澤 史子

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中 1 0 1 5 番地 富士通株式会社内

【氏名】 富川 真弓

【特許出願人】

【識別番号】 000005223

【氏名又は名称】 富士通株式会社

【代表者】 関澤 義

【代理人】

【識別番号】 100087848

【弁理士】

【氏名又は名称】 小笠原 吉義

【電話番号】 03-3823-2606

## 【手数料の表示】

【納付方法】 予納

【予納台帳番号】 012586

【納付金額】 14,000円

## 【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9001635

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 立体構造処理装置

【特許請求の範囲】

【請求項 1】 立体構造を形成する二つの順序付けられていない点集合または二つの順序付けられた点集合または二つの部分的に対応付けられた点集合のデータを計算機により処理し、これらの点集合のデータによって表される二つの立体構造を重ね合わせる立体構造処理装置であって、

前記二つの点集合の各要素に対応付ける際に、幾何学的な関係による候補の絞り込み、所定のしきい値条件による候補の絞り込み、または点の属性による候補の絞り込みを行い、これらの幾何学的制約、しきい値条件または点の属性を満たす要素の組合せを生成する組合せ生成処理手段(14)と、

生成された組合せの中で各点間の距離の平均が小さくなる点集合間の対応付けを探し出すことによって二つの立体構造の最もよく一致する重ね合わせのための位置と方向を算出する重ね合わせ算出処理手段(15)と、

算出された重ね合わせの結果を出力する結果出力処理手段(18)とを備えたことを特徴とする立体構造処理装置。

【請求項 2】 物質の二つの立体構造に関するデータを計算機により処理し、これらの二つの立体構造を重ね合わせる立体構造処理装置であって、

物質の立体構造に関するデータが登録されたデータベース(21)と、

登録されているデータおよびユーザからの指示情報を読み込むデータ入力処理手段(17)と、

読み込んだ二つの立体構造に関するデータに基づき、二つの立体構造を形成する点集合の各要素に対応付ける際に、幾何学的な関係による候補の絞り込み、所定のしきい値条件による候補の絞り込み、または点の属性による候補の絞り込みを行い、これらの幾何学的制約、しきい値条件または点の属性を満たす要素の組合せを生成する組合せ生成処理手段(14)と、

生成された組合せの中で各点間の距離の平均が小さくなる点集合間の対応付けを探し出すことによって二つの立体構造の最もよく一致する重ね合わせのための位置と方向を算出する重ね合わせ算出処理手段(15)と、

算出された結果に基づいて物質の立体構造同士を重ね合わせて表示する結果出力処理手段(18)とを備えた

ことを特徴とする立体構造処理装置。

【請求項3】 タンパク質を構成する物質の二つの立体構造に関するデータを計算機により処理し、これらの二つの立体構造を重ね合わせる立体構造処理装置であって、

タンパク質の立体構造に関するデータが登録されたデータベース(21)と、

登録されているデータおよびユーザからの指示情報を読み込むデータ入力処理手段(17)と、

読み込んだタンパク質の二つの立体構造に関するデータに基づき、アミノ酸配列番号によりタンパク質を構成する各アミノ酸を順序集合として捉え、それらの各要素を対応付ける際に、幾何学的な関係による候補の絞り込み、所定のしきい値条件による候補の絞り込み、またはアミノ酸の属性による候補の絞り込みを行い、これらの幾何学的制約、しきい値条件またはアミノ酸の属性を満たす要素の組合せを生成する組合せ生成処理手段(14)と、

生成された組合せの中で各要素間の距離の平均が小さくなるアミノ酸の順序集合間の対応付けを探し出すことによって、タンパク質の二つの立体構造の最もよく一致する重ね合わせのための位置と方向を算出する重ね合わせ算出処理手段(15)と、

算出された結果に基づいてタンパク質の立体構造同士を重ね合わせて表示する結果出力処理手段(18)とを備えた

ことを特徴とする立体構造処理装置。

【請求項4】 物質の立体構造に関するデータが格納されたデータベース(21)を備え、指定された物質の立体構造を検索キーとして、前記データベース(21)を検索することにより、該データベース(21)中から類似した立体構造を抽出する立体構造処理装置であって、

前記データベース(21)に登録されているデータおよびユーザからの指示情報を読み込むデータ入力処理手段(17)と、

検索キーとなる指定された物質の立体構造と前記データベース(21)から読み込

んだ立体構造に関するデータに基づき、類似性判定のために二つの立体構造を形成する点集合の各要素を対応付ける際に、幾何学的な関係による候補の絞り込み、所定のしきい値条件による候補の絞り込み、または点の属性による候補の絞り込みを行い、これらの幾何学的制約、しきい値条件または点の属性を満たす要素の組合せを生成する組合せ生成処理手段(14)と、

生成された組合せの中で各点間の距離の平均が小さくなる点集合間の対応付けを探し出すことによって、前記データベース(21)中から類似性の高い立体構造を選び出す類似性算出処理手段(16)と、

選び出された類似性の高い立体構造を検索結果として出力する結果出力処理手段(18)とを備えた

ことを特徴とする立体構造処理装置。

【請求項5】 タンパク質を構成する物質の立体構造に関するデータが格納されたデータベース(21)を備え、指定された立体構造を検索キーとして前記データベース(21)を検索することにより、該データベース(21)中から類似した立体構造を抽出する立体構造処理装置であって、

前記データベース(21)に登録されているデータおよびユーザからの指示情報を読み込むデータ入力処理手段(17)と、

検索キーとなる指定された立体構造と前記データベース(21)から読み込んだタンパク質の立体構造に関するデータに基づき、類似性の判定のためにそれらの各要素を対応付ける際に、アミノ酸配列番号によりタンパク質を構成する各アミノ酸を順序集合として捉え、幾何学的な関係による候補の絞り込み、所定のしきい値条件による候補の絞り込み、またはアミノ酸の属性による候補の絞り込みを行い、これらの幾何学的制約、しきい値条件またはアミノ酸の属性を満たす要素の組合せを生成する組合せ生成処理手段(14)と、

生成された組合せの中で各要素間の距離の平均が小さくなるアミノ酸の順序集合間の対応付けを探し出すことによって、前記データベース(21)中から類似性の高い立体構造を選び出す類似性算出処理手段(16)と、

選び出された類似性の高い立体構造を検索結果として出力する結果出力処理手段(18)とを備えた

ことを特徴とする立体構造処理装置。

【請求項 6】 物質の立体構造に関するデータが格納されたデータベース (21) を備え、指定されたある機能を発現する立体構造をもとに、それに類似する機能を発現する立体構造を前記データベース (21) 中から抽出して収集する立体構造処理装置であって、

前記データベース (21) に登録されているデータおよびユーザからの指示情報を読み込むデータ入力処理手段 (17) と、

指定されたある機能を発現する立体構造と前記データベース (21) から読み込んだ立体構造に関するデータに基づき、類似性判定のために二つの立体構造を形成する点集合の各要素を対応付ける際に、幾何学的な関係による候補の絞り込み、所定のしきい値条件による候補の絞り込み、または点の属性による候補の絞り込みを行い、これらの幾何学的制約、しきい値条件または点の属性を満たす要素の組合せを生成する組合せ生成処理手段 (14) と、

生成された組合せの中で各点間の距離の平均が小さくなる点集合間の対応付けを探し出すことによって、前記データベース (21) 中から類似性の高い立体構造を選び出す類似性算出処理手段 (16) と、

選び出された類似性の高い立体構造を収集し、その収集した立体構造に関するデータを、指定された機能に対応させて格納することにより機能データベース (24) を生成する機能データベース生成処理手段 (19) とを備えた

ことを特徴とする立体構造処理装置。

【請求項 7】 タンパク質を構成する物質の立体構造に関するデータが格納されたデータベース (21) を備え、指定されたある機能を発現する立体構造をもとに、それに類似する機能を発現する立体構造を前記データベース (21) 中から抽出して収集する立体構造処理装置であって、

前記データベース (21) に登録されているデータおよびユーザからの指示情報を読み込むデータ入力処理手段 (17) と、

指定されたある機能を発現する立体構造と前記データベース (21) から読み込んだタンパク質の立体構造に関するデータに基づき、類似性判定のためにそれらの各要素を対応付ける際に、アミノ酸配列番号によりタンパク質を構成する各アミ



ノ酸を順序集合として捉え、幾何学的な関係による候補の絞り込み、所定のしきい値条件による候補の絞り込み、またはアミノ酸の属性による候補の絞り込みを行い、これらの幾何学的制約、しきい値条件またはアミノ酸の属性を満たす要素の組合せを生成する組合せ生成処理手段(14)と、

生成された組合せの中で各要素間の距離の平均が小さくなるアミノ酸の順序集合間の対応付けを探し出すことによって、前記データベース(21)中から類似性の高い立体構造を選び出す類似性算出処理手段(16)と、

選び出された類似性の高いタンパク質の機能に関連する立体構造を収集し、その収集した立体構造に関するデータを、指定された機能に対応させて格納することにより機能データベース(24)を生成する機能データベース生成処理手段(19)とを備えた

ことを特徴とする立体構造処理装置。

【請求項 8】 物質の機能と関連する立体構造に関するデータが格納された機能データベース(24)を備え、指定された物質の立体構造をもとに、その立体構造が発現する機能を予測する立体構造処理装置であって、

物質を構成する立体構造のデータを読み込むデータ入力処理手段(17)と、

前記機能データベース(24)に登録されている立体構造と指定された立体構造に関するデータに基づき、類似性判定のためそれらの二つの立体構造を形成する点集合の各要素を対応付ける際に、幾何学的な関係による候補の絞り込み、所定のしきい値条件による候補の絞り込み、または点の属性による候補の絞り込みを行い、これらの幾何学的制約、しきい値条件または点の属性を満たす要素の組合せを生成する組合せ生成処理手段(14)と、

生成された組合せの中で各点間の距離の平均が小さくなる点集合間の対応付けを探し出すことによって、類似性の高い立体構造を前記機能データベース(24)中から選び出す類似性算出処理手段(16)と、

選び出された類似性の高い立体構造が発現する機能を、指定された物質の立体構造が発現する機能であるとして予測し、その結果を出力する機能予測処理手段(20)とを備えた

ことを特徴とする立体構造処理装置。

【請求項 9】 タンパク質の機能と関連する立体構造に関するデータが格納された機能データベース(24)を備え、指定されたあるタンパク質を構成する物質の立体構造をもとに、その立体構造が発現する機能を予測する立体構造処理装置であって、

前記機能データベース(24)に登録されているタンパク質の立体構造に関するデータおよびユーザからの指示情報を読み込むデータ入力処理手段(17)と、

前記機能データベース(24)に登録されている立体構造と指定されたタンパク質の立体構造に関するデータに基づき、類似性判定のためそれらの各要素を対応付ける際に、アミノ酸配列番号によりタンパク質を構成する各アミノ酸を順序集合として捉え、幾何学的な関係による候補の絞り込み、所定のしきい値条件による候補の絞り込み、またはアミノ酸の属性による候補の絞り込みを行い、これらの幾何学的制約、しきい値条件またはアミノ酸の属性を満たす要素の組合せを生成する組合せ生成処理手段(14)と、

生成された組合せの中で各要素間の距離の平均が小さくなるアミノ酸の順序集合間の対応付けを探し出すことによって、類似性の高い立体構造を前記機能データベース(24)の中から選び出す類似性算出処理手段(16)と、

選び出された類似性の高い立体構造が発現する機能を、指定されたタンパク質の立体構造が発現する機能であるとして予測し、その機能および機能部位に関する情報を出力する機能予測処理手段(20)とを備えた

ことを特徴とする立体構造処理装置。

【請求項 1 0】 立体構造に関するデータが格納されたデータベース(21)を備え、指定された立体構造を検索キーとして、前記データベース(21)を検索することにより、該データベース(21)の中から類似した立体構造を抽出する立体構造処理装置であって、

前記データベース(21)に登録されているデータおよびユーザからの指示情報を読み込むデータ入力処理手段(17)と、

検索キーとなる立体構造および検索対象となる立体構造の各々を構成する順序付けられていない点集合または順序付けられた点集合または部分的に対応付けられた点集合の各要素を対応付ける際に、各点集合を構造的に意味のある部分構造

に対応する点の部分集合に分割する部分構造算出処理手段(12)と、

検索キー側の部分構造と検索対象側の部分構造との類似性を調べ、部分構造が類似している場合に限って点集合全体の対応付けを行い、各点間の距離の平均が小さくなる点集合間の対応付けを探し出すことによって、前記データベース(21)中から類似性の高い立体構造を選び出す類似性算出処理手段(16)と、

選び出された類似性の高い立体構造を検索結果として出力する結果出力処理手段(18)とを備えた

ことを特徴とする立体構造処理装置。

【請求項 1 1】 分子の立体構造に関するデータが格納されたデータベース(21)を備え、指定された立体構造を検索キーとして、前記データベース(21)を検索することにより、該データベース(21)中から類似した分子の立体構造を抽出する立体構造処理装置であって、

前記データベース(21)に登録されているデータおよびユーザからの指示情報を読み込むデータ入力処理手段(17)と、

読み込んだ分子の立体構造に関するデータを構造的に意味のある部分構造に対応する座標の部分集合に分割する部分構造算出処理手段(12)と、

算出された部分構造の名前とそれを構成する原子の座標とが格納される部分構造座標テーブル(13)と、

該部分構造座標テーブル(13)中のデータに基づき、検索キー側の部分構造と検索対象側の部分構造との類似性を調べ、部分構造が類似している場合に限って原子を要素とする集合全体の対応付けを行い、各要素間の距離の平均が小さくなる集合間の対応付けを探し出すことによって、前記データベース(21)中から類似性の高い立体構造を選び出す類似性算出処理手段(16)と、

選び出された類似性の高い分子の立体構造を検索結果として出力する結果出力処理手段(18)とを備えた

ことを特徴とする立体構造処理装置。

【請求項 1 2】 タンパク質を構成する物質の立体構造に関するデータが格納されたデータベース(21)を備え、指定された立体構造を検索キーとして前記データベース(21)を検索することにより、該データベース(21)中から類似した立体

構造を抽出する立体構造処理装置であって、

読み込んだタンパク質の立体構造に関するデータを部分構造の一種である二次構造に対応する座標の部分集合に分割する部分構造算出処理手段(12)と、

算出された二次構造の名前とそれを構成するアミノ酸の座標とが格納される部分構造座標テーブル(13)と、

該部分構造座標テーブル(13)中のデータに基づき、検索キー側の部分構造と検索対象側の部分構造との類似性を調べ、部分構造が類似している場合に限りアミノ酸を要素とする集合全体の対応付けを行い、各要素間の距離の平均が小さくなる集合間の対応付けを探し出すことによって、前記データベース(21)中から類似性の高い立体構造を選び出す類似性算出処理手段(16)と、

選び出された類似性の高いタンパク質の立体構造を検索結果として出力する結果出力処理手段(18)とを備えた

ことを特徴とする立体構造処理装置。

【請求項 1 3】 分子の二つの立体構造に関するデータを計算機により処理し、これらの二つの立体構造を重ね合わせる立体構造処理装置であって、

分子の立体構造に関するデータが登録されたデータベース(21)と、

登録されているデータおよびユーザからの指示情報を読み込むデータ入力処理手段(17)と、

読み込んだ分子の立体構造に関するデータを構造的に意味のある部分構造に対応する座標の部分集合に分割する部分構造算出処理手段(12)と、

算出された部分構造の名前とそれを構成する原子の座標とが格納される部分構造座標テーブル(13)と、

該部分構造座標テーブル(13)中のデータに基づき、各部分構造について相手側の部分構造との類似性を調べ、部分構造が類似している場合に限り原子を要素とする集合全体の対応付けを行い、各要素間の距離の平均が小さくなる集合間の対応付けを探し出すことによって二つの立体構造の最もよく一致する重ね合わせのための位置と方向を算出する重ね合わせ算出処理手段(15)と、

算出された結果に基づいて分子の立体構造同士を重ね合わせて表示する結果出力処理手段(18)とを備えた

ことを特徴とする立体構造処理装置。

【請求項14】 タンパク質を構成する物質の二つの立体構造に関するデータを計算機により処理し、これらの二つの立体構造を重ね合わせる立体構造処理装置であって、

タンパク質の立体構造に関するデータが登録されたデータベース(21)と、登録されているデータおよびユーザからの指示情報を読み込むデータ入力処理手段(17)と、

読み込んだタンパク質の立体構造に関するデータを部分構造の一種である二次構造に対応する座標の部分集合に分割する部分構造算出処理手段(12)と、

算出された二次構造の名前とそれを構成するアミノ酸の座標とが格納される部分構造座標テーブル(13)と、

該部分構造座標テーブル(13)中のデータに基づき、各部分構造について相手側の部分構造との類似性を調べ、部分構造が類似している場合に限ってアミノ酸を要素とする集合全体の対応付けを行い、各要素間の距離の平均が小さくなる集合間の対応付けを探し出すことによって、二つの立体構造の最もよく一致する重ね合わせのための位置と方向を算出する重ね合わせ算出処理手段(15)と、

算出された結果に基づいて立体構造同士を重ね合わせて表示する結果出力処理手段(18)とを備えた

ことを特徴とする立体構造処理装置。

【請求項15】 立体構造を形成する二つの順序付けられた点集合A、Bのデータを計算機により処理し、これらの点集合のデータによって表される二つの立体構造を重ね合わせる立体構造処理装置であって、

前記二つの点集合A、Bの各要素を対応付ける際に、点集合Aの構成要素数に応じて、対応付けの候補となる点の数を所定の関数 $f(x)$ によって設定し、他方の点集合Bの先頭から所定数 $\Delta f(x)$ ずつ点を選んで対応付けの候補とし、幾何学的な関係による候補の絞り込み、所定のしきい値条件による候補の絞り込み、または点の属性による候補の絞り込みを行い、これらの幾何学的制約、しきい値条件または点の属性を満たす要素の組合せを生成する組合せ生成処理手段(14)と、

生成された組合せの中で各点間の距離の平均が小さくなる点集合間の対応付けを探し出すことによって二つの立体構造の最もよく一致する重ね合わせのための位置と方向を算出する重ね合わせ算出処理手段(15)と、

算出された重ね合わせの結果を出力する結果出力処理手段(18)とを備えたことを特徴とする立体構造処理装置。

【請求項 1 6】 立体構造を形成する二つの順序付けられた点集合 A、B のデータを計算機により処理し、これらの点集合のデータによって表される二つの立体構造を重ね合わせる立体構造処理装置であって、

前記二つの点集合 A、B の各要素を対応付ける際に、点集合 A の両端の二点間の距離を  $d$  とし、点集合 B において距離が  $d \pm \alpha$  (ただし、 $\alpha$  は所定の正数) にある二点を選び、その二点の間にある点を対応付けの候補とし、幾何学的な関係による候補の絞り込み、所定のしきい値条件による候補の絞り込み、または点の属性による候補の絞り込みを行い、これらの幾何学的制約、しきい値条件または点の属性を満たす要素の組合せを生成するとともに、次の組合せの決定においては既に対応付けられた点の位置または点集合 A の要素数をもとに次の対応付けの候補を決定する組合せ生成処理手段(14)と、

生成された組合せの中で各点間の距離の平均が小さくなる点集合間の対応付けを探し出すことによって二つの立体構造の最もよく一致する重ね合わせのための位置と方向を算出する重ね合わせ算出処理手段(15)と、

算出された重ね合わせの結果を出力する結果出力処理手段(18)とを備えたことを特徴とする立体構造処理装置。

【請求項 1 7】 請求項 1 5 または請求項 1 6 記載の立体構造処理装置において、

前記立体構造は、タンパク質を構成する物質の立体構造であり、

前記順序付けられた点集合 A、B は、タンパク質を構成するアミノ酸配列順序に基づく各アミノ酸の順序集合であることを特徴とする立体構造処理装置。

【請求項 1 8】 タンパク質を構成する物質の立体構造に関するデータが格納されたデータベース(21)を備え、指定された立体構造を検索キーとして前記データベース(21)を検索することにより、該データベース(21)の中から類似した立体

構造を抽出する立体構造処理装置であって、

前記データベース(21)に登録されているデータおよびユーザからの指示情報を読み込むデータ入力処理手段(17)と、

検索キーとなる指定された立体構造と前記データベース(21)から読み込んだタンパク質の立体構造に関するデータに基づき、類似性の判定のためにそれらの各要素を対応付ける際に、アミノ酸配列番号によりタンパク質を構成する各アミノ酸を順序集合として捉え、検索キー側の順序集合Aのアミノ酸数に応じて、対応付けの候補となるアミノ酸の数を所定の関数  $f(x)$  によって設定し、検索対象側の順序集合Bの先頭から所定数  $\Delta f(x)$  ずつアミノ酸を選んで対応付けの候補とし、幾何学的な関係による候補の絞り込み、所定のしきい値条件による候補の絞り込み、またはアミノ酸の属性による候補の絞り込みを行い、これらの幾何学的制約、しきい値条件またはアミノ酸の属性を満たす要素の組合せを生成する組合せ生成処理手段(14)と、

生成された組合せの中で各要素間の距離の平均が小さくなるアミノ酸の順序集合間の対応付けを探し出すことによって、前記データベース(21)中から類似性の高い立体構造を選び出す類似性算出処理手段(16)と、

選び出された類似性の高い立体構造を検索結果として出力する結果出力処理手段(18)とを備えた

ことを特徴とする立体構造処理装置。

【請求項19】 タンパク質を構成する物質の立体構造に関するデータが格納されたデータベース(21)を備え、指定された立体構造を検索キーとして前記データベース(21)を検索することにより、該データベース(21)中から類似した立体構造を抽出する立体構造処理装置であって、

前記データベース(21)に登録されているデータおよびユーザからの指示情報を読み込むデータ入力処理手段(17)と、

検索キーとなる指定された立体構造と前記データベース(21)から読み込んだタンパク質の立体構造に関するデータに基づき、類似性の判定のためにそれらの各要素を対応付ける際に、アミノ酸配列番号によりタンパク質を構成する各アミノ酸を順序集合として捉え、検索キー側の順序集合Aの両端のアミノ酸間の距離を

dとし、検索対象側の順序集合Bにおいて距離が $d \pm \alpha$ （ただし、 $\alpha$ は所定の正数）にあるアミノ酸を選び、そのアミノ酸間にあるアミノ酸を対応付けの候補とし、幾何学的な関係による候補の絞り込み、所定のしきい値条件による候補の絞り込み、またはアミノ酸の属性による候補の絞り込みを行い、これらの幾何学的制約、しきい値条件またはアミノ酸の属性を満たす要素の組合せを生成するとともに、次の組合せの決定においては既に対応付けられたアミノ酸の位置または順序集合Aの要素数をもとに次の対応付けの候補を決定する組合せ生成処理手段(14)と、

生成された組合せの中で各要素間の距離の平均が小さくなるアミノ酸の順序集合間の対応付けを探し出すことによって、前記データベース(21)中から類似性の高い立体構造を選び出す類似性算出処理手段(16)と、

選び出された類似性の高い立体構造を検索結果として出力する結果出力処理手段(18)とを備えた

ことを特徴とする立体構造処理装置。

#### 【発明の詳細な説明】

#### 【0001】

#### 【産業上の利用分野】

本発明は、物質の二つの立体構造を重ね合わせたり、類似部分を抽出したりする立体構造処理装置に関するものである。

#### 【0002】

物理・化学の分野では、新しい（未知の）物質の性質を調べたり、新しい物質を人工的に創製するために、分子構造を分析し、物質が持つ機能の発現メカニズムを解明する研究が行われている。これまでの研究成果により、物質の持つ機能と立体構造との間には密接な関係があることが知られており、構造的に類似した部分（あるいは特異的な部分）が物質の機能に大きく関与すると考えられている。このため、X線結晶解析装置やNMRなどによって物質の立体構造を決定し、その結果明らかになった立体構造のデータベース化が図られている。

#### 【0003】

例えば、このデータベースから立体構造中の類似部分を計算機で自動的に抽出



したり、検索したりすることができれば、従来研究者が行ってきた一連の作業を軽減することができる。

#### 【0004】

##### 【従来の技術】

物理・化学の分野では、新しい（未知の）物質の性質を調べたり、新しい物質を人工的に創製するために、X線結晶解析装置やNMR等の手法で物質の立体構造を決定し、決定された立体構造の情報をデータベースに蓄積することが行われている。代表的なデータベースとして、タンパク質のX線結晶解析により明らかになったタンパク質等の立体構造を登録したプロテイン・データ・バンク（PDB: Protein Data Bank）が広く知られており、世界的に用いられている。また、化学物質が登録されているデータベースとして、ケンブリッジ・ストラクチャー・データベース（CSD: Cambridge Structural Database）が知られている。

#### 【0005】

タンパク質は複数のアミノ酸が一本の鎖のように連結し、この鎖が生体内で折りたたまることによって立体構造を形成し、各種の機能を発現するようになっている。各アミノ酸はN末端からC末端に向けて1から順に番号付けることによって表現される。これらの番号はアミノ酸番号、アミノ酸配列番号またはアミノ酸残基番号などと呼ばれている。また、各アミノ酸はその種類に応じて複数の原子より構成される。したがって、上記PDBには、タンパク質の名前、管理番号、タンパク質を形成するアミノ酸番号、各アミノ酸を構成する各原子の種類と三次元座標等の情報が登録されている。

#### 【0006】

これまでの化学的な研究成果から、物質の立体構造とその機能との間には密接な関係があることが知られており、物質の改変や新しい機能を持つ物質を創製するために、化学的な実験を通じて立体構造と機能の関係が解明されつつある。その中でも、同じ機能を持つ物質間で構造的に類似した部分（あるいは特異的な部分）が物質の機能に大きく関与すると考えられるため、立体構造中に共通に存在する類似な構造を探し出すことは必要不可欠である。

#### 【0007】

しかし、三次元座標から直接特徴的な部分を取り出す手法がないため、研究者は、各立体構造を三次元グラフィックシステムで表示し、それを見て手で特徴的な部分を探しているのが現状である。一般に物質には定まった向きの決め方がなく、片方の物質を基準としてもう一方の物質を回転させながら特徴的な部分を探し出すため、これらの作業にはかなりの時間を要している。

#### 【0008】

研究者が類似する立体構造を探す際、物質の立体構造の類似性の尺度として  $r.m.s.d.$  (root mean square distance) 値が使用されている。 $r.m.s.d.$  値は互いに対応付けられた物質の構成要素間の平均二乗距離の平方根を表す値である。経験的には、物質間の  $r.m.s.d.$  値が  $1 \text{ \AA}$  以下の場合に、それらの物質はきわめて類似していると考えられている。

#### 【0009】

図2に  $r.m.s.d.$  値の算出方法を示す。

例えば図2の(a)に示すような点集合  $A = \{a_1, a_2, \dots, a_i, \dots, a_m\}$  で表される物質と、(b)に示すような点集合  $B = \{b_1, b_2, \dots, b_j, \dots, b_n\}$  で表される物質があったとする。これらの物質AとBを構成する要素を(c)に示すように互に対応付け、対応付けられた要素間の  $r.m.s.d.$  値が最小になるように物質Bを回転、移動させていく。 $r.m.s.d.$  値は、対応付ける点の数を  $n$ 、回転行列を  $U$ 、各重みを  $w_n$  とすると、次式の

$$r.m.s.d. = \sqrt{(\sum w_k (U b_k - a_k)^2) \div n}$$

(ただし、 $k$  は  $1$  から  $n$  まで総和)

で求められる。なお、この対応付けられた点同士間で  $r.m.s.d.$  値を最小化する物質の回転、移動を求める手法はカブシュ等により提案され、現在広く利用されている。しかし、本手法は同数の点同士を比較するため、一方の物質中のどの構成要素を他方の物質のどの構成要素に対応付けると  $r.m.s.d.$  値が最小になるかについて、研究者が試行錯誤しながら求めているのが現状である。

#### 【0010】

また、新しい物質を創製するためには、既存の物質を調べる必要がある。例えばある物質の耐熱性を強化したい場合には、耐熱性の強い物質に共通する構造を

探し出し、そのような構造を新たに作成する物質に付加することによって機能の強化を図ったりする。これには、データベースの中から必要な構造を検索する機能が必要になるが、前述の理由より同様にコンピュータグラフィックシステムを用いて研究者が試行錯誤しながら必要な構造をデータベースから探し出しているのが現状である。

#### 【0011】

以上のように、従来、作業者は解析したい物質の立体構造をグラフィックシステムを用いてグラフィカルに表示し、画面上で他の分子との視覚的な比較、重ね合わせなどの操作により手作業で解析を行わなければならなかった。

#### 【0012】

また、タンパク質の立体構造中には共通してみられる基本的な構造が存在し、これらを二次構造と呼んでいる。r.m.s.d.を使わず二次構造の類似性により自動的に検索する方法についても、現在までにいくつか研究されているが、それらはアミノ酸配列順序に沿って部分構造を二次構造の記号で表し、その記号によって比較する方法であり、空間的な位置関係の類似性による比較はできなかった。

#### 【0013】

##### 【発明が解決しようとする課題】

以上のように従来、CSDやPDBを利用して物質の立体構造を解析する場合、大量のデータからの構造的な検索、構造比較を手作業でするために多くの時間や労力を要し、作業者の大きな負担となっていた。また、そのためにデータベースのデータを有効に利用することができず、物質の構造解析を十分にできないといった問題点があった。したがって、立体構造データベースの立体構造の類似性による検索システムが必要とされていた。

#### 【0014】

本発明は上記問題点の解決を図り、立体構造の類似した部分構造を自動的に抽出することによって複数の立体構造の重ね合わせ表示や、データベースからの類似構造の検索を自動化し、化学物質やタンパク質の構造解析における構造比較や、類似構造の検索の作業を効率化し、作業に要する時間や作業者の負担を軽減し、また従来ではできなかった全データを対象とした立体構造の解析を可能とする

システムを提供することを目的とする。

【0015】

すなわち、本発明は、立体構造中の類似部分を計算機で自動的に抽出することによってコンピュータグラフィックシステムにおける立体構造の重ね合わせ表示の自動化を可能とすることを目的とする。

【0016】

また、本発明は、データベースから類似な立体構造を検索することを可能とし、各作業に要する時間、人員等のコストを削減するとともに、各作業を効率化する物質の解析・検索システムを提供することを目的とする。

【0017】

また、本発明は、物質の立体構造が格納されたデータベースから、ある既知の機能を発現する立体構造と類似するものを探し出すことによって、機能データベースを自動生成する手段を提供することを目的とする。

【0018】

また、本発明は、物質の機能と関連する立体構造に関するデータが格納された機能データベースをもとに、ある立体構造の未知の機能を予測する手段を提供することを目的とする。

【0019】

さらに、本発明は、以上の処理を高速化し、実用上十分な短時間で処理するシステムを提供することを目的とする。

【0020】

【課題を解決するための手段】

図1は本発明の原理説明図である。

図1において、10はCPUおよびメモリなどからなる処理装置、11はオペレーティング・システムなどの制御部、12は物質の立体構造のデータをその部分構造に対応する座標の部分集合に分割する部分構造算出処理手段、13は部分構造名とその要素の位置を記憶する部分構造座標テーブル、14は二つの集合の各要素を対応付ける組合せを所定の制約条件のもとに生成する組合せ生成処理手段、15は二つの物質等を重ね合わせる重ね合わせ算出処理手段、16は二つの

物質の類似性を調べる類似性算出処理手段、17は物質等のデータを入力するデータ入力処理手段、18は検索結果等を出力する結果出力処理手段、19は機能データベースを自動生成する機能データベース生成処理手段、20は指定された物質の未知の機能を予測する機能予測処理手段、21は物質の立体構造に関するデータが格納されたデータベース、22はキーボードやポインティング・デバイス等の入力装置、23は表示装置、24は機能データベースを表す。

#### 【0021】

組合せ生成処理手段14は、タンパク質や各種分子等の立体構造を構成する要素に関する二つの集合の各要素を対応付ける際に、幾何学的な関係による候補の絞り込み、所定のしきい値条件による候補の絞り込み、または点の属性による候補の絞り込みを行い、これらの幾何学的制約、しきい値条件または点の属性を満たす要素の組合せを生成する処理手段である。

#### 【0022】

重ね合わせ算出処理手段15は、組合せ生成処理手段14により生成された組合せの中で各要素間の距離の平均が小さくなる要素集合間の対応付けを探し出すことによって二つの立体構造の最もよく一致する重ね合わせのための位置と方向を算出する処理手段である。

#### 【0023】

結果出力処理手段18は、算出された重ね合わせの結果または後述する検索結果などを表示装置23に出力する処理手段である。

データベース21には、タンパク質などの物質の立体構造に関するデータが登録されている。データ入力処理手段17は、データベース21に登録されているデータを読み込んだり、また入力装置22からユーザが入力したコマンド等の指示情報を読み込んだりする処理手段である。

#### 【0024】

類似性算出処理手段16は、指定された物質の立体構造を検索キーとし、データベース21中の類似した立体構造を検索するために、組合せ生成処理手段14が生成した二つの立体構造の要素の組合せについて、空間的に類似するものを探し出す処理手段である。

**【 0 0 2 5 】**

機能データベース生成処理手段 1 9 は、指定されたある機能を発現する立体構造をもとに、それに類似する機能を発現する立体構造をデータベース 2 1 中から抽出して収集するために、類似性算出処理手段 1 6 が選び出した高い類似性を持つ立体構造を収集し、その収集した立体構造に関するデータを機能別に格納することにより機能データベース 2 4 を生成する処理手段である。

**【 0 0 2 6 】**

機能予測処理手段 2 0 は、指定された機能が未知の物質の立体構造と機能データベース 2 4 に登録されている機能が既知の立体構造とを照合し、組合せ生成処理手段 1 4 および類似性算出処理手段 1 6 の処理により選び出された類似性の高い立体構造が発現する機能を、指定された物質の立体構造が発現する機能であるとして予測し、その結果を出力する処理手段である。

**【 0 0 2 7 】**

部分構造算出処理手段 1 2 は、立体構造の検索にあたって、検索キーとなる立体構造および検索対象となる立体構造の各々を構成する要素からなる集合の各要素を対応付ける際に、各集合を構造的に意味のある部分構造に対応する要素の部分集合に分割することにより、処理の高速化を図る処理手段である。

**【 0 0 2 8 】**

部分構造テーブル 1 3 は、部分構造算出処理手段 1 2 により分割された情報が設定される記憶手段である。

この場合、類似性算出処理手段 1 6 は、検索キー側の部分構造と検索対象側の部分構造との類似性を調べ、部分構造が類似している場合に限って集合全体の対応付けを行い、各要素間の距離の平均が小さくなる点集合間の対応付けを探し出すことによって、データベース 2 1 中から類似性の高い立体構造を選び出す。

**【 0 0 2 9 】**

また、組合せ生成処理手段 1 4 は、タンパク質などの立体構造において、類似構造の抽出、検索を行う際に、または二つの立体構造を重ね合わせる際に、検索のキーとなる構造の形状や大きさに応じて探索空間を分割し、対応付けの候補となる要素の組合せを狭い探索空間内で生成することにより、計算時間の短縮を可

能とする。

#### 【0030】

##### 【作用】

本発明によれば、立体構造を形成する点集合、順序関係のある点集合および部分的に対応付けられた点集合同士をr.m.s.d.値が最適な値をとるように重ね合わせるができるため、従来人手に頼っていた立体構造の重ね合わせ作業を自動化することができ、作業の効率化を図ることができる。

#### 【0031】

また、本発明によれば、タンパク質の立体構造データベースより、類似性の高い構造を探し出すことを自動化することにより、作業者の負担を軽減することができる。従来のように、グラフィカルに表示して他のタンパク質と比較する手間を減らすことができる。また、作業を自動化することにより、従来ではできなかった全データを対象とした立体構造の解析を行うことが可能になる。

#### 【0032】

また、機能データベースの自動生成や、機能データベースを用いることによる機能予測の自動化が可能になる。

さらに、請求項10ないし請求項14記載の発明によれば、各立体構造を部分構造に分割し、それらの部分構造の類似性を考慮して点の対応付けを行うので、処理の高速化が可能になる。

#### 【0033】

また、請求項15ないし請求項19記載の発明によれば、点同士を対応付ける際に、検索キーとなる構造の形状や大きさに応じて探索空間を分割し処理を行うので、短時間で点（要素）同士を対応付けることができるようになる。

#### 【0034】

##### 【実施例】

以下、図面を参照しつつ本発明の実施例を説明する。

最初に、図1に示す組合せ生成処理手段14により点を対応付ける基本的な処理について説明する。

〔組合せ生成処理手段14による対応付けの処理の説明〕

## (1) 点集合の対応付け

物質Aと物質Bが各々点集合  $A = \{a_1, a_2, \dots, a_i, \dots, a_m\}$ ,  $1 \leq i \leq m$ ,  $B = \{b_1, b_2, \dots, b_j, \dots, b_n\}$ ,  $1 \leq j \leq n$  で形成され、各点は  $a_i = (x_i, y_i, z_i)$ ,  $b_j = (x_j, y_j, z_j)$  からなる三次元座標であるとする。この場合、各点集合の要素の対応付けは原理的には各集合内の点を順次対応付けていき、図3の(A)に示すような木構造を作成することによってすべての組合せを生成することができる。

## 【0035】

図3の(B)は、点集合Aの要素が3個、点集合Bの集合が4個の場合の対応付けの例、すなわち点集合  $A = \{a_1, a_2, a_3\}$  と  $B = \{b_1, b_2, b_3, b_4\}$  の対応付けを表している。点線は生成された候補を表し、その中で最適な対応付け ( $a_1$  と  $b_2$ ,  $a_2$  と  $b_3$ ,  $a_3$  と  $b_4$ ) を実線で表している。

## 【0036】

図中の  $nil$  は、対応する点が存在しない場合に相当する。 $nil$  を対応付けることによって比較する点集合の要素の数が異なる場合でも、最適な対応付けを生成することができる。このようにして生成された組合せの中から、平均二乗距離の平方根の値 (r.m.s.d. 値) が最小になる組合せを選択することにより最良の対応付けを生成することができる。

## 【0037】

図4に順序付けられていない点の要素からなる点集合AとBの対応付けアルゴリズムの例を示す。

点集合Aから1つずつ要素aを取り出し、まだ木構造中の先祖、兄弟に含まれていない要素bと組み合わせて、それが制約条件を満たすかどうかを調べる。制約条件を満たしていれば、それを木構造中に登録し、次の要素についての対応付けを行っていく。

## 【0038】

しかし、このような手法を取ると、例えばnのm乗個の組合せが生成されるため、一般には計算不可能な問題となってしまう。具体的には、順序づけられていない点集合A (m点) と点集合B (n点) の場合、生成される組合せの数は、i



を  $n i l$  の数とすると,

【0 0 3 9】

【数 1】

$$\sum_{i=0}^m {}_n P_{m-i} \times {}_m C_i = \sum_{i=0}^m \frac{n!}{(n-m+i)!} \times \frac{m!}{i!(m-i)!}$$

【0 0 4 0】

となる。ここで,  $n=4$ ,  $m=3$  とすると,

【0 0 4 1】

【数 2】

$$\begin{aligned} \sum_{i=0}^3 {}_4 P_{3-i} \times {}_3 C_i &= \sum_{i=0}^3 \frac{4!}{(4-3+i)!} \times \frac{3!}{i!(3-i)!} \\ &= \frac{4!}{1!} \times \frac{3!}{3!} + \frac{4!}{2!} \times \frac{3!}{1!2!} + \frac{4!}{3!} \times \frac{3!}{2!1!} + \frac{4!}{4!} \times \frac{3!}{3!} \\ &= 24 + 36 + 12 + 1 = 73 \end{aligned}$$

【0 0 4 2】

となる。すなわち, 図 3 (B) に示す点集合 A (3 点) と点集合 B (4 点) のような場合, 生成される組合せの数は 73 通りとなる。実際には, これらの点 (要素) の数はもっとずっと多いのが普通であるので, 組合せの数は膨大な数となる。

【0 0 4 3】

そこで, これらの対応付けを行う際に, 以下の (4), (5), (6) に示す各点集合内の幾何学的な関係, しきい値条件, 点の属性を考慮することによって最適な組合せを生成することを図る。

【0 0 4 4】

(2) 順序付けられた点集合の対応付け

物質 A と物質 B が各々点集合  $A = \{a_1, a_2, \dots, a_i, \dots, a_m\}$ ,  $1 \leq i \leq m$ ,  $B = \{b_1, b_2, \dots, b_j, \dots, b_n\}$ ,  $1 \leq j \leq n$  で形成され, 各点は  $a_i = (x_i, y_i, z_i)$ ,  $b_j = (x_j, y_j, z_j)$  からなる三次元

座標であり、点集合Aでは、 $a_1 < a_2 < \dots < a_i < \dots < a_m$  (または  $a_1 > a_2 > \dots > a_i > \dots > a_m$ ) なる順序関係が成り立ち、同様に点集合B内でも  $b_1 < b_2 < \dots < b_j < \dots < b_n$  (または  $b_1 > b_2 > \dots > b_j > \dots > b_n$ ) なる順序関係が成り立つとする。

#### 【0045】

この場合、各点集合の要素の対応付けは原理的には各集合内の点を順序関係に基づいて順次対応付けていき、図5 (A) に示す木構造を作成することによってすべての組合せを生成することができる。図5 (B) は、点集合Aの要素が3点、点集合Bの要素が4点の場合の例である。すなわち図5 (B) は、順序付けられた点集合  $A = \{a_1, a_2, a_3\}$  , (順序関係は  $a_1 < a_2 < a_3$ ) と、同様に順序付けられた点集合  $B = \{b_1, b_2, b_3, b_4\}$  , (順序関係は  $b_1 < b_2 < b_3 < b_4$ ) の対応付けを表している。

#### 【0046】

点線は生成された候補を表し、その中で最適な対応付け ( $a_1$  と  $b_2$  ,  $a_2$  と  $b_3$  ,  $a_3$  と  $b_4$ ) を実線で表している。図中の  $n_{il}$  は対応する点が存在しない場合に相当する。 $n_{il}$  を対応付けることによって比較する点集合の要素の数が異なる場合でも最適な対応付けを生成することができる。このようにして生成された組合せの中から、 $r.m.s.d.$  値が最小になる組合せを選択することにより、最良の対応付けを生成することができる。

#### 【0047】

また、点集合内の各要素に対して順序関係を導入することによって上記(1)と比べて大幅に生成する組合せの数を減らすことができる。さらに、これらの対応付けを行う際に、後述する(4) , (5) , (6) に示す各点集合内の幾何学的な関係、しきい値、点の属性を考慮することによって、最適な組合せを生成することができる。

#### 【0048】

図6に順序付けられた点集合の対応付けアルゴリズムの例を示す。

点集合Aから1つずつ要素aを取り出し、まだ木構造中の先祖、兄弟に含まれていない要素の中で親ノードの要素より大きい要素b<sub>j</sub>と組み合わせて、それが

制約条件を満たすかどうかを調べる。制約条件を満たしていれば、それを木構造中に登録し、次の要素についての対応付けを行っていく。

【0049】

このように順序付けられた点集合の場合、生成される組合せの数は、

【0050】

【数3】

$$\sum_{i=0}^m {}_nC_{m-i} \times {}_mC_i = \sum_{i=0}^m \frac{n!}{(m-i)!(n-m+i)!} \times \frac{m!}{i!(m-i)!}$$

【0051】

となる。ここで、 $n=4$ 、 $m=3$ とすると、

【0052】

【数4】

$$\begin{aligned} \sum_{i=0}^3 {}_4C_{3-i} \times {}_3C_i &= \sum_{i=0}^3 \frac{4!}{(3-i)!(4-3+i)!} \times \frac{3!}{i!(3-i)!} \\ &= \frac{4!}{3!1!} \times \frac{3!}{3!} + \frac{4!}{2!2!} \times \frac{3!}{1!2!} + \frac{4!}{1!3!} \times \frac{3!}{2!1!} + \frac{4!}{4!} \times \frac{3!}{3!} \\ &= 4 + 18 + 12 + 1 = 35 \end{aligned}$$

【0053】

となり、図5(B)のような点集合A(3点)と点集合B(4点)の場合には、生成される組合せの数は35通りとなる。

(3) 部分的に対応付けられた点集合または順序付けられた点集合の対応付け

上記(1)または(2)の場合において、部分的に対応付けられる点同士が決まっていることがある。この場合、各点集合の要素の対応付けは、予め部分的に対応が付けられている点の情報を参照しながら、残りの点に対して(1)、(2)の手法と同様に各集合内の点を順次対応付けていき、図7に示すような木構造を作成することによってすべての組合せを生成することができる。

【0054】

図7中の×印は部分的な対応付けによって枝刈りされる部分を表す。図では予

め点集合Aの要素 $a_1$ と点集合Bの要素 $b_2$ とが対応付けられていた場合の対応付けを表している。また、(1)、(2)と同様にこれらの対応付けを行う際に、後述する(4)、(5)、(6)に示す各点集合内の幾何学的な関係、しきい値条件、点の属性を考慮することによって最適な組合せを生成することができる。

#### 【0055】

##### (4) 幾何学的な関係による候補の絞り込み

幾何学的関係に基づいて点集合の要素を対応付けることによって、無駄な組合せを生成することを防ぐことができるため、効率よく点集合を対応付けることができる。

#### 【0056】

##### (a) 距離関係による候補の絞り込み

対応付けを行う際に、点集合A内の要素 $a_i$ と近接する $s$ 個( $1 \leq s \leq m-1$ ,  $n-1$ )の点間に成り立つ距離関係 $|a_i - a_{i-s}|$ と、点集合B内の要素 $b_j$ と近接する $s$ 個の要素間の距離関係 $|b_j - b_{j-s}|$ とが、許容誤差範囲 $\Delta d$ ( $||a_i - a_{i-s}| - |b_j - b_{j-s}|| \leq \Delta d$ )の関係にある点のみを選択し、対応付けることによって対応付ける候補数を絞り込むことができる。

#### 【0057】

図8に点集合Aの要素 $a_i$ に対応する点集合B内の点 $b_j$ を選択する場合の幾何学的関係の使用例を示す。図中の数値は距離を表す。

図8の(A1)、(A2)に示すように、点集合A内の要素 $a_i$ と近接する $s=2$ 個の点 $a_{i-1}$ 、 $a_{i-2}$ 間に成り立つ距離関係 $|a_i - a_{i-1}| = 2.0$ 、 $|a_i - a_{i-2}| = 3.0$ に対して、点集合B内の要素 $b_p$ 、 $b_q$ 、 $b_r$ の中で各点と隣接する2個の要素間の距離関係 $|b_j - b_{j-s}|$ が許容誤差範囲 $\Delta d = 0.5$ である点のみを選択し、対応付ける。本図では幾何学的関係として点同士の相互距離を比較した結果、点 $b_p$ ( $|b_p - b_{j-1}| = 2.2$ 、 $|b_p - b_{j-2}| = 3.3$ )が距離関係を満たすため、点 $b_p$ が $b_j$ の候補として選択される。

#### 【0058】

##### (b) 角度による候補の絞り込み

立体構造が類似している場合には、立体構造を形成する各点間の角度の関係も

類似していると考えることができる。立体では、図8の(B1)に示すように3点間の角度 $\theta$ と4点間のうち各3点から形成される面間の角度 $\phi$ が存在する。以下では、3点間の角度 $\theta$ を例として対応付ける点を絞り込む方法について述べる。

#### 【0059】

対応付けを行う際に、点集合A内の要素 $a_i$ と近接する $s$ 個( $2 \leq s \leq m-1$ ,  $n-1$ )の点間に成り立つ角度に対し、点集合B内の要素 $b_j$ と近接する $s$ 個の要素間の角度が許容誤差範囲 $\Delta\theta$ である点のみを選択し、対応付けることによって対応付ける候補数を絞り込む。

#### 【0060】

図8(B2)に、点集合Aの間に成り立つ幾何学的な関係として各要素間の角度を考え、それに基づいて点集合Bの点を対応付ける場合の例を示す。

点集合A内の要素 $a_i$ と近接する $s=2$ 個の点 $a_{i-1}$ ,  $a_{i-2}$ 間に成り立つ角度が $\theta_a$ , 点集合B内の要素 $b_p$ ,  $b_q$ ,  $b_r$ の中で各点と隣接する2個の要素 $b_{j-1}$ ,  $b_{j-2}$ 間の角度が各々 $\theta_{b1}$ ,  $\theta_{b2}$ ,  $\theta_{b3}$ の場合、角度差が許容誤差範囲 $\Delta\theta$ である点のみを選択し、それらに対応付ける。本図では点 $b_q$ に関してのみ許容誤差範囲 $|\theta_a - \theta_{b1}| \leq \Delta\theta$ を満たすため、 $b_j$ の候補として $b_q$ が選択される。

#### 【0061】

(c) 重心からの距離、角度による候補の絞り込み

立体構造が類似している場合には、重心からの距離、角度も類似していると考えられる。従って、選択した点間で重心を算出し、上記(a), (b)と同様の手法で距離、角度を比較することによって、対応付ける候補を絞り込むことができる。

#### 【0062】

(5) しきい値条件による候補の絞り込み

上述の(1)～(4)の方法において、所定のしきい値を設定し、候補の持つ属性値がしきい値より大きい値を持つならば探索の枝刈りを行うことによって、対応付けの効率化を図ることができる。このしきい値として、例えば $n$ 個の制限やr.m.s.d.値の制限を用いることができる。

## 【0063】

## (a) nil 数の制限

生成した組合せの中で nil の総数が多くなり過ぎると、結果的に意味のない組合せの候補が生成される。そこで、点集合 A と点集合 B の各要素間の対応付けを行う際に、nil の総数があるしきい値以上になったならば、それを組合せの候補から除外することによって無駄な候補の生成を避け、効率よく対応付けを行うことができる。

## 【0064】

図9は、点集合  $A = \{a_1, a_2, a_3\}$ 、 $B = \{b_1, b_2, b_3, b_4\}$  を対応付ける際に nil の総数を0個に制限した場合の枝刈りの例を示す。図中、木構造中で×印が付けられた部分が、枝刈りされる部分である。

## 【0065】

## (b) r.m.s.d. 値の制限

点集合 A の要素  $a_i$  に点集合 B の要素  $b_j$  を対応付けることによって、これまでに対応付けられた全点間の r.m.s.d. 値が極端に悪くなる場合には、その点を候補から除外することが望ましい。そこで、要素  $a_i$  に要素  $b_j$  を対応付けた場合の全点間の r.m.s.d. 値を算出し、r.m.s.d. 値があるしきい値以下ならばその点を候補とし、そうでなければ候補から除外することによって対応付ける点の候補を効率よく生成することができる。

## 【0066】

## (6) 点の属性による候補の絞り込み

点集合 A の要素  $a_i$  に点集合 B の要素  $b_j$  を対応付ける際に、点の属性を利用することによって対応付ける点の候補を絞り込むことができる。点の属性として、例えば原子、原子団、分子の種類や、親水性、疎水性や、電荷の正負などを挙げることができる。これらが一致するかどうかを調べることによって、候補に加えるかどうかを判定する。

## 【0067】

例えば、タンパク質の構成要素間に対応付ける場合には、点の属性としてアミノ酸残基の種類（原子団に相当する）を用いることにより、対応付ける候補を絞

り込むことができる。なお、アミノ酸残基の種類等に関しては、例えば「『生化学の基礎』（東京化学同人出版）p.21～26」等の参考資料を参照されたい。

#### 【0068】

また、ある特定の要素に対して制限を加えることによって、同様に対応付ける点の候補を絞ることができる。例えば、ある点に対しては `nil` を挿入しないという制限を設けたり、ある点に対して点の属性を指定することなどにより、検索するものに対してよりきめの細かい絞り込みが可能になる。

#### 【0069】

次に、物質の立体構造としてタンパク質を題材とした例を説明する。ただし、対象は基本的には立体座標であれば特に限定されないため、一般の分子構造についても同様の手法で適用することができる。

〔請求項1ないし請求項9記載の発明の実施例〕

##### (1) 分子構造を重ね合わせる処理装置の例

物質の性質を調べるときには、各分子同士を重ね合わせてお互いに共通な部分や特異的な部分を判別することによって、各物質の性質を分析したり、あるいは予測したりすることができる。これらの作業は、従来人手で行われているため、各分子構造を自動的に重ね合わせて表示する装置が必要とされている。

#### 【0070】

それを実現する分子構造重ね合わせ表示装置のシステムは、図1に示す装置構成のうち、物質の立体構造に関する情報を登録したデータベース21、登録されたデータおよびユーザからの入力コマンドを読み込むデータ入力処理手段17、重ね合わせる立体構造の要素を対応付ける組合せ生成処理手段14、データベース21から読み込んだ物質の立体構造（三次元座標）に基づいて、`r.m.s.d.` 値が最小となるように立体構造同士を重ね合わせる重ね合わせ算出処理手段15および算出された結果に基づいて立体構造同士を重ね合わせて表示する結果出力処理手段18等を用いて構成される。

#### 【0071】

##### (a) データベース21

物質の立体構造に関する情報を格納したデータベース21である。データベー

ス 2 1 には、物質の名称、物質を構成する原子の三次元座標等が格納されている。

#### 【 0 0 7 2 】

##### (b) データ入力処理手段 1 7

データ入力処理手段 1 7 では、ユーザの入力コマンドに基づいて重ね合わせる物質のデータ（三次元座標）をデータベース 2 1 から読み込み、組合せ生成処理手段 1 4、重ね合わせ算出処理手段 1 5 へ送る処理を行う。

#### 【 0 0 7 3 】

##### (c) 組合せ生成処理手段 1 4

各要素を対応付ける際に、幾何学的な関係による候補の絞り込み、所定のしきい値条件による候補の絞り込み、または点の属性による候補の絞り込みを行い、これらの幾何学的制約、しきい値条件または点の属性を満たす要素の組合せを生成する。

#### 【 0 0 7 4 】

ここでの対応付けでは、タンパク質を構成するアミノ酸配列順序に基づいて空間的に類似な部分に対応付ける機能とアミノ酸配列順序とは無関係に空間的に類似な部分に対応付ける機能を提供する。アミノ酸配列順序に基づいて空間的に類似な部分を検索する場合には、タンパク質を構成する各アミノ酸をアミノ酸配列番号により順序付けられた順序集合として捉えることができ、対応付け処理の説明における(2)～(6)で述べた方式で類似な部分を算出することができる。また、各アミノ酸を単なる集合として捉えることによって、上記(1)、(3)～(6)で述べた方式により、アミノ酸配列順序とは無関係に空間的に類似な部分を算出することができる。

#### 【 0 0 7 5 】

##### (d) 重ね合わせ算出処理手段 1 5

重ね合わせ算出処理手段 1 5 では、物質の立体構造（三次元座標）について、r.m.s.d. 値が最適な値をとるように物質を構成する要素間の対応付けを行い、その座標変換情報を求めて、結果を結果出力処理手段 1 8 へ送る。

#### 【 0 0 7 6 】



## (e) 結果出力処理手段18

結果出力処理手段18によるグラフィック表示では、重ね合わせ算出処理手段15で算出された結果に基づいて物質の立体構造を重ね合わせて表示する。表示結果を回転させながら見るができるようにすることによって、どの部分がどのように重なっているかを三次元グラフィックにより判別することを、さらに容易化することができる。

## 【0077】

タンパク質Calmodulinのアミノ酸残基配列を図10(A)に、TroponinCのアミノ酸残基配列を図10(B)に示す。図10(C)はTroponinCの立体構造を示している。なお、図10(A)、(B)は、プロテイン・データ・バンク(PDB)に登録されているアミノ酸残基配列を抜粋したものである。

## 【0078】

図10(A)に示したアミノ酸残基配列は、本来のアミノ酸残基配列に対してアミノ酸配列番号1-4、148番に相当するアミノ酸が欠落しているため番号がずれているが、以降の説明では、図に示したアミノ酸配列番号を使用することにする。

## 【0079】

Calmodulinは生化学的な実験結果より、4個の $Ca^{2+}$ を結合することが知られている。 $Ca^{2+}$ の結合部位はアミノ酸配列中に4箇所(サイト)あり、その中でアミノ酸配列番号81-108、117-143は、TroponinC中に2箇所ある $Ca^{2+}$ の結合部位と各々似た骨格を取ることが知られている。タンパク質はアミノ酸で構成されるが、その骨格は各アミノ酸を構成する一つの原子( $C\alpha$ )の座標で代表できることが知られている。

## 【0080】

そこで、Calmodulinの $Ca^{2+}$ の結合部位81-108を検索キー(検索キーをプローブともいう)として、アミノ酸配列順序に基づいて空間的に類似な部分(単一サイト)を検索した結果を、図11に示す。図11の結果から、Calmodulinの $Ca^{2+}$ の結合部位81-108に対応付けられたTroponinC中のアミノ酸配列番号は96-123番であることがわかる。本結果は、生化学的な実験結果と一

致するものである。

#### 【0081】

図12にCalmodulinのCa<sup>2+</sup>の結合部位81-108と117-143とをブロープとしてアミノ酸配列順序に基づいて空間的に類似な部分(複数サイト)を検索した結果を示す。図12から、CalmodulinのCa<sup>2+</sup>の結合部位81-108、117-143に対応付けられたTroponinC中のアミノ酸配列番号は、各々96-123、132-158であることがわかる。本結果も同様に生化学的な実験結果と一致するものである。

#### 【0082】

このように、本装置を使用すれば、物質間の立体構造のr.m.s.d.値が小さくなるような物質の構成要素間の対応付けがなされるため、対応付けられた部分を重ね合わせて表示することによって、物質の最適な重ね合わせ表示を実現することができる。

#### 【0083】

##### (2) 立体構造検索装置、機能データベース生成装置の例

新薬の開発のように新しい機能を持つ物質を開発したり、既に存在する物質の機能強化を図るためには、物質の機能と構造の相関関係を解明することが不可欠である。このような作業を進めるに当たって、類似な立体構造を持つ物質を多数参照する必要がある。そのためには、データベースから立体構造が類似した物質を簡単に取り出せる立体構造検索装置が必要になる(また、このような装置があると機能と関係する立体構造を収集した機能データベースを作成することができる。機能データベースについては次の(3)で後述する)。

#### 【0084】

それを実現する立体構造検索装置は、図1に示す装置構成のうち、物質の立体構造に関する情報を登録したデータベース21、登録されたデータおよびユーザからの入力コマンドを読み込むデータ入力処理手段17、比較する立体構造の要素を対応づける組合せ生成処理手段14、データベース21から読み込んだ物質の立体構造(三次元座標)について、r.m.s.d.値が最小となる類似な構造を検索する類似性算出処理手段16、および検索結果を表示する結果出力処理手段18

等を用いて構成される。

【0085】

同様に機能データベース生成装置は、データベース 21、データ入力処理手段 17、組合せ生成処理手段 14、類似性算出処理手段 16 および機能データベース生成処理手段 19 等を用いて構成される。

【0086】

(a) データベース 21

物質の立体構造に関する情報を格納したデータベースである。データベースには、物質の名称、物質を構成する原子の三次元座標等が格納されている。

【0087】

(b) データ入力処理手段 17

立体構造のデータおよびユーザからの入力コマンドに基づいて検索のキーとなる立体構造および検索時に参照するデータベース 21 に登録されている立体構造のデータを読み込み、組合せ生成処理手段 14 および類似性算出処理手段 16 へ送る。

【0088】

(c) 組合せ生成処理手段 14 および類似性算出処理手段 16

類似性判定のために各要素を対応付ける際に、幾何学的な関係による候補の絞り込み、所定のしきい値条件による候補の絞り込み、または点の属性による候補の絞り込みを行い、これらの幾何学的制約、しきい値条件または点の属性を満たす要素の組合せを生成する。この際、タンパク質を構成するアミノ酸配列順序に基づいて空間的に類似な部分を検索する機能とアミノ酸配列順序とは無関係に空間的に類似な部分を検索する機能を提供する。アミノ酸配列順序に基づいて空間的に類似な部分を検索する場合には、タンパク質を構成する各アミノ酸をアミノ酸配列番号により順序付けられた順序集合として捉えることができ、対応付け処理の説明における(2)～(6)で述べた方式で類似な部分を算出することができる。また、各アミノ酸を単なる集合として捉えることによって、上記(1)、(3)～(6)で述べた方式により、アミノ酸配列順序とは無関係に空間的に類似な部分を算出することができる。これにより類似性算出処理手段 16 は、類似性の高い立

体構造をデータベース 2 1 中から選出する。

【 0 0 8 9 】

(d) 結果出力処理手段 1 8

結果出力処理手段 1 8 では、類似性算出処理手段 1 6 の結果に基づいて類似な部分をアミノ酸配列名およびアミノ酸番号で表し、類似性の尺度として r.m.s.d. 値を表示する。

【 0 0 9 0 】

(e) 機能データベース生成処理手段 1 9

また、機能データベース生成装置の場合、機能データベース生成処理手段 1 9 は、類似性算出処理手段 1 6 の結果に基づいて、プローブの立体構造と類似する立体構造を、プローブの立体構造が発現する機能と同じ機能を持つものとする。そして、その情報を機能データベース 2 4 に格納する。

【 0 0 9 1 】

タンパク質 elongation factor の G T P ( グアノシン 3 リン酸 ) のリン酸結合部位であるアミノ酸残基番号 7 から 1 4 に対応する C  $\alpha$  の座標をプローブとして P D B から類似な立体構造の検索を行った結果を、図 1 3 に示す。

【 0 0 9 2 】

この例では、プロテイン・データ・バンク ( P D B ) に登録されているデータ 9 0 5 件中でタンパク質の立体構造 7 4 4 件を検索の対象とした。検索結果として、図 1 3 に示すように、検索したターゲットタンパク質のアミノ酸残基番号、アミノ酸残基配列、プローブのアミノ酸残基配列およびターゲットとプローブの立体構造の r.m.s.d. 値等の情報が出力される。途中の出力を図示省略するが、検索の結果、8 個の立体構造が検索されている ( プローブ自身を含む ) 。

【 0 0 9 3 】

タンパク質の種類別にみると、adenylate kinase が 3 件、elongation factor が 2 件 ( 内 1 件はプローブ自身 )、ras protein が 3 件であり、いずれも A T P または G T P のリン酸結合部位であった。これから、A T P または G T P のリン酸との結合という機能と立体構造は非常に密接な関係にあり、かつ他のリン酸結合部位とは無関係な構造と偶然一致することがないため非常に特異的な構造を持

っていることがわかる。

#### 【0094】

このように、本装置を使用すれば、プローブ（検索キー）となる物質の立体構造を指定することによって、物質の立体構造を格納したデータベース 21 から類似構造を検索することができる。また、この検索結果を利用し、機能データベース 24 を自動生成することができる。

#### 【0095】

##### (3) 機能予測装置

図 13 に示した結果から推察されるように、タンパク質は、ある機能を発現するためにその機能に特異的な立体構造を持っていると考えられている。したがって、機能ごとにその機能に特異的な立体構造のデータベース（図 1 に示す機能データベース 24）があると、新しく X 線結晶解析や NMR 等の手法で物質の立体構造が決定された際に、その機能データベース 24 に登録されている構造が立体構造内にあるかを調べることによって、その物質がどのような機能を持っており、その機能は立体構造中のどの部分（以下機能部位とよぶ）によって司られているかを予測することができる。

#### 【0096】

それを実現する機能予測装置は、図 1 に示す装置構成のうち、データ入力処理手段 17、機能と関連した立体構造を登録した機能データベース 24、これから読み込んだ立体構造の要素と、比較する立体構造の要素とを対応付ける組合せ生成処理手段 14、類似性を調べる類似性算出処理手段 16、その結果から機能と機能部位を予測する機能予測処理手段 20 等を用いて構成される。

#### 【0097】

##### (a) データ入力処理手段 17

物質を構成する立体構造のデータを読み込み、組合せ生成処理手段 14 へ送る。

#### 【0098】

##### (b) 機能データベース 24

物質の機能とその機能に特異的な立体構造に関する情報を格納したデータベー

スである。機能データベース 2 4 には、機能の名称、機能に特異的な立体構造を構成する原子の三次元座標等が格納されている。

#### 【0 0 9 9】

##### (c) 組合せ生成処理手段 1 4 および類似性算出処理手段 1 6

ここでは、機能データベース 2 4 に登録されている立体構造と入力された立体構造間の最適な重ね合わせを算出する。この際、タンパク質を構成するアミノ酸配列順序に基づいて空間的に類似な部分を検索する機能とアミノ酸配列順序とは無関係に空間的に類似な部分を検索する機能を提供する。アミノ酸配列順序に基づいて空間的に類似な部分を検索する場合には、タンパク質を構成する各アミノ酸をアミノ酸配列番号により順序付けられた順序集合として捉えることができ、対応付け処理の説明における(2)～(6)で述べた方式で類似な部分を算出することができる。また、各アミノ酸を単なる集合として捉えることによって、上記(1)、(3)～(6)で述べた方式により、アミノ酸配列順序とは無関係に空間的に類似な部分を算出することができる。

#### 【0 1 0 0】

##### (d) 機能予測処理手段 2 0

機能予測処理手段 2 0 では、類似性算出処理手段 1 6 の結果に基づいて機能データベース 2 4 に登録されている機能名、機能部位のアミノ酸配列名およびアミノ酸残基番号で表し、類似性の尺度として r.m.s.d. 値を表示する。

【請求項 1 0 ないし請求項 1 4 記載の発明の実施例】

化学物質やタンパク質の構造解析における類似構造の検索では、2つの立体構造を比較して類似性を算出しなければならない。そのために形成する点間の対応付けを自動化することが必要になる。本実施例では、立体構造を形成する点間の対応付けを容易にかつ高速化するために、点集合を構造的に意味のある部分構造に対応する点の部分集合に分割し、相手側の部分構造との類似性を調べ、部分構造が類似している場合に限って点集合全体の対応付けを行う。

#### 【0 1 0 1】

以下に構成要素の対応付けの具体的な方法を述べる。

##### (1) 点集合の部分集合への分割

構造Aは点集合  $A = \{a_1, a_2, a_3, \dots, a_i, \dots, a_m\}$ ,  $1 \leq i \leq m$ ,  $B = \{b_1, b_2, b_3, \dots, b_j, \dots, b_n\}$ ,  $1 \leq j \leq n$  から構成され、各点は  $a_i = (x_i, y_i, z_i)$ ,  $b_j = (x_j, y_j, z_j)$  からなる三次元座標で表されとする。

#### 【0102】

各点間の対応付けを容易にするために、構造を構造的に意味のある部分構造に分割し、点集合を部分集合に分割する。構造的に意味のある部分構造の例として、化学物質では、官能基やある機能をもった部分構造、タンパク質では二次構造であるヘリックスやシート構造、機能を発現するために必要な部分構造などが挙げられる。

#### 【0103】

部分構造の座標は既知のデータを使用したり、あるいは三次元座標より解析することによって求める。部分集合に分割された点集合Aを  $A = \{a_1, a_2, \dots, a_k\}$ ,  $\{a_{k+1}, a_{k+2}, \dots, a_l\}$ ,  $\dots$ ,  $\{a_{l+1}, a_{l+2}, \dots, a_m\}$  とし、 $1 \leq k \leq l \leq m$  とする。ここで、 $SA_1 = \{a_1, a_2, \dots, a_k\}$ ,  $SA_2 = \{a_{k+1}, a_{k+2}, \dots, a_l\}$ ,  $\dots$ ,  $SA_p = \{a_{l+1}, a_{l+2}, \dots, a_m\}$  とすると、その集合  $SA$  は点集合Aを構成する部分集合であり、AはSAによって  $A = \{SA_1, SA_2, \dots, SA_p\}$  と表される。同様に点集合BもBの部分集合であるSBに分割することによって  $B = \{SB_1, SB_2, \dots, SB_q\}$  と表される。

#### 【0104】

##### (2) 部分集合間の対応付け

構造A =  $\{SA_1, SA_2, \dots, SA_p\}$  と構造B =  $\{SB_1, SB_2, \dots, SB_q\}$  の各要素を対応付けることを考える。この場合、要素間の対応付けは各集合内の要素を順次対応付けていき、対応付けた状態を表した木構造を作成することにより記述することができる。木のルート(root)の節点は開始点である。葉節点は可能な対応付けの結果を表し、中間節点は部分的な結果を表す。対応する要素が存在しない場合には nil を対応付ける。

#### 【0105】

全ての組み合わせに対応する状態木を生成すると、節点の数が膨大になるため、枝刈りすることを考える。2つの部分集合を対応付けて節点を加える際に部分集合間でマッチングを行い、その結果が制約条件を満たすものであれば節点を加えることにする。制約条件については後記の(4)で述べる。部分集合間のマッチングに関しては、上述した実施例による立体構造を形成する順序付けられた2つの点集合の各要素を対応付ける立体構造のマッチング方法によって行う。

#### 【0106】

部分集合の対応付けを表す図を、図14に示す。

(3) 部分的に対応付けられた部分集合または順序付けられた部分集合の対応付け

上記(2)の場合において部分的に対応付けられる部分集合同士が決まっている場合、対応付けられている部分集合を節点を生成する候補から取り除き木構造を生成することによって対応付ける。この場合にも(2)と同様の枝刈りを行う。

#### 【0107】

(4) 部分集合間の類似性による候補の絞り込み

上記(2)、(3)の方法において、候補である2つの部分集合間の類似性により枝刈りし、効率よく対応付けを行うことができる。類似性として、候補の持つ属性と2つの構造的な類似性を考える。部分集合の属性は化学物質では官能基の種類や機能の種類、タンパク質では二次構造の構成要素や機能の種類などがある。2つの部分集合の構造的な類似性を判定するためには、前の実施例で説明した立体構造検索装置の立体構造を形成する順序付けられた2つの点集合の各要素を対応付ける立体構造のマッチング方法を用いる。そして、この方法による最適なマッチングを行ったときに得られる各点間の平均二乗距離の平方根(r.m.s.d.)を算出する。

#### 【0108】

候補となる2つの部分集合の属性が同じで、かつr.m.s.d.値がしきい値以下の場合だけそれらに対応付けた節点を生成することにより、候補を絞り込むことができる。以上の制約条件を考慮した集合AとBの部分集合の対応付けアルゴリズムを、図15に示す。



## 【0109】

点集合Aより部分集合を取り出し、それをSAとする。また、点集合SBより木構造中の先祖、兄弟に含まれない要素SBを取り出し $d_j$ とする。取り出せる要素がない場合は $d_j = nil$ とする。

## 【0110】

そして、SAと $d_j$ の属性が同じであるかどうかを調べ、同じでなければその組合せを棄却し、枝刈りする。属性が同じであれば、各点集合のマッチングを行い、最適なマッチングでのr.m.s.d.値を算出する。これが所定のしきい値以内であれば、SAと $d_j$ を対応付け木構造中の $d_{j-1}$ の子ノードとして登録し、最適な点の対応付けを配列に格納する。以上の処理をすべての部分集合について繰り返す。

## 【0111】

以上の方法を用いて実現可能なタンパク質立体構造の二次構造的類似性による検索システムについて述べる。

## (1) タンパク質立体構造の二次構造的類似性による検索システム

図16にその検索処理機構の構成を示す。本装置は、タンパク質の立体構造情報が登録されているデータベース160（図1に示すデータベース21に相当するもの）、データベース中の立体構造情報から二次構造を決定し部分構造に分割する二次構造算出部161（図1に示す部分構造算出処理手段12に相当するもの）、二次構造算出部161によって得られた結果を二次構造のタイプと構成する点の三次元座標として格納する二次構造座標テーブル162（図1に示す部分構造座標テーブル13に相当するもの）、ユーザの入力コマンドを読み込む入力部163（図1に示すデータ入力処理手段17に相当するもの）、入力されたコマンドと二次構造座標テーブルのデータに基づいて、前述した手法により類似構造を検索する検索部164（図1に示す類似性算出処理手段16に相当するもの）、および検索結果をグラフィカルに表示する表示部165（図1に示す結果出力処理手段18に相当するもの）から構成される。以下に各部の詳細を述べる。

## 【0112】

## (a) データベース160

タンパク質立体構造情報が格納されているデータベースである。タンパク質ごとに名前と構成する原子の三次元座標データが登録されている。タンパク質の立体構造は一本の鎖状のアミノ酸からなり、構成するアミノ酸は順序付けられている。各アミノ酸の位置をアミノ酸の構成原子の一つである  $C_{\alpha}$  の座標で代表することとし、順序付けられたアミノ酸の配列の一方の端を 1 番目とすると  $n$  個のアミノ酸からなるタンパク質は順序付けられた  $n$  個の座標の点集合として表される。  $i$  番目 ( $1 \leq i \leq n$ ) のアミノ酸の座標は  $i$  番目の座標に対応する。

#### 【0113】

##### (b) 二次構造算出部 161

データベースの三次元座標に基づいてタンパク質の構造を二次構造のタイプに分割し、点集合を部分集合に分割する。二次構造のタイプとその定義の例を、図 17 に示す。この図 17 に示す定義に従って  $i = 1$  から順に  $i$  番目のアミノ酸がどのタイプに属するかを決定し、一連の座標のグループを一つの部分集合とする。決定された二次構造のタイプと構成するアミノ酸の座標データとを二次構造座標テーブル 162 に格納する。この作業を繰り返し  $n$  個のアミノ酸全てを部分集合に分割する。以上の二次構造の決定と部分集合への分割に関する処理の流れを、図 18 に示す。

#### 【0114】

##### (c) 二次構造座標テーブル 162

二次構造座標テーブル 162 の構成例を、図 19 に示す。これには、二次構造算出部 161 により決定された二次構造のタイプと二次構造を構成するアミノ酸の座標データが格納されている。この例では、部分集合  $S_1$ ,  $S_2$  は、 $\alpha$  ヘリックスのタイプ、部分集合  $S_3$ , ... は、 $\beta$  シートのタイプに属している。

#### 【0115】

##### (d) 入力部 163

入力部 163 では、二次構造座標テーブル 162 およびユーザからの入力コマンドに基づいて検索キーとなるタンパク質名を読み込む。それを検索部 164 に送る。

#### 【0116】

## (e) 検索部 1 6 4

検索部 1 6 4 では、入力部 1 6 3 から送られてきたキーとなるタンパク質の二次構造座標テーブル 1 6 2 に格納されているデータを読み込み、部分集合の対応付けを行い、2つの構造間の平均二乗距離の平方根 (r.m.s.d.) を算出し、r.m.s.d. 値がしきい値よりも小さいものを選ぶことによって類似性の高い構造を検索する。対応付けの際に前述した部分集合間の対応付けの方法を用いる。この場合、部分集合の属性は二次構造のタイプである。対応付けの候補である2つの部分集合のタイプが一致し、構造を最適にマッチングさせた時のr.m.s.d. 値がしきい値以下の場合にのみ対応付けを行う。

## 【0 1 1 7】

次に、対応付けられた点同士をマッチングさせ、構造全体のr.m.s.d. 値を算出する。r.m.s.d. 値がしきい値以下の場合類似構造として検索結果に登録する。この作業を二次構造座標テーブル 1 6 2 に格納されたタンパク質の全てに関して行い全てのデータから二次構造的に類似した立体構造を検索する。

## 【0 1 1 8】

以上の検索部 1 6 4 での処理の例を、図 2 0 に示す。

## (f) 表示部 1 6 5

検索部 1 6 4 で得られた検索結果に基づいて、類似構造を持つタンパク質名、およびキータンパク質と類似構造タンパク質の二次構造と、二次構造を構成するアミノ酸を表示する。

## 【0 1 1 9】

図 2 1 に、その出力例を示す。また、図 2 2 に、検索したキータンパク質 A とそれに類似する構造を持つ検索結果であるタンパク質 B の立体構造を示す。

図 2 2 では、 $\alpha$  ヘリックスの部分構造をリボンのらせんで、 $\beta$  ストランドの部分構造を矢印で、ループやターンの部分構造をチューブで示している。この結果より、キータンパク質はアミノ酸配列の順に  $\alpha$  ヘリックス、 $\beta$  ストランド、ループ、 $\beta$  ストランドの4つの部分構造に分割され、それぞれ部分集合 S A 1, S A 2, S A 3, S A 4 に対応していることがわかる。

## 【0 1 2 0】

図21に示すように、A中のSA1, SA2, SA4はそれぞれB中の矢印で示されている部分構造SB10, SB1, SB3と類似し、3つの部分構造の空間的位置関係も類似している。A中のループ部分SA3に矢印がないのは類似する部分構造がなかったことを示す。類似構造タンパク質B中の類似部分は、図22(B)の図においてハッチングを施した部分である。

〔請求項15ないし請求項19記載の発明の実施例〕

最初の実施例で説明した立体構造の検索システムでは、立体構造を形成する点集合間の要素の対応付けを、各集合内の点を順次対応付けていくことにより自動化した。この手法では、対応付けるべき点がない場合にnilに対応付けることによって、比較する点集合の要素の数が異なる場合でも、最適な対応付けを生成することができるようにしている。このようにして、生成された組合せの中からr.m.s.d.値が最小になる組合せを選択することにより、最良の対応付けを生成する。

#### 【0121】

ここで、組合せに関する探索の枝刈りが行われない場合、それぞれの点集合の構成要素数をm, n個とすると、生成される組合せ数は $m^n$ 個となり、実質的に計算不可能な問題となってしまう。そこで、これらの対応付けを行う際に、各点集合の要素間の距離などの幾何学的関係、r.m.s.d.値、nilの個数などのしきい値条件、また、構成要素の属性（タンパク質の場合、アミノ酸の種類など）を考慮することによって最適な組合せを生成するようにしている。どのような条件下でも高速に処理する手法を確立することは困難であるが、場合により、さらに計算時間の高速化を図ることができることがある。

#### 【0122】

本実施例では、類似構造の抽出、検索のキーとなる立体構造の形状に注目し、検索などの処理の一層の高速化を可能とする。

図23に本実施例を適用する対象を説明するための直線的構造と非直線的構造の例を示す。

#### 【0123】

図23(A)に示すように、立体構造の両端の2点が最も遠い距離にあるよう

な構造を直線的構造と呼ぶ。一方、図23(B)に示すように、両端の2点が最も遠い距離ではない構造を非直線的構造と呼ぶ。本実施例は、図23(A)に示すような直線的構造を持つ立体構造に対して適用できる方式であり、さらに高速な検索処理等を可能とするものである。

#### 【0124】

本実施例では、立体構造を形成する点集合A、B間の要素の対応付けを行う際、点集合Aの空間的な大きさや構成要素の数などに応じて、点集合Bを分割し、対応付けの候補となる点の部分集合を求めてから対応付けを行うことによって、効率良く組合せを生成する。以下において、対応付けの候補となる部分集合を求める手段について述べる。

#### 【0125】

(1) 点集合Aの構成要素数に応じた順序付けられた点集合Bの分割

図24は、点集合Aの構成要素数に応じた順序づけられた点集合Bの分割の例を説明するための図である。

#### 【0126】

点集合Aの要素数 $m$ に応じた探索空間の大きさを設定し、その大きさを順序付けられた点集合Bを分割し探索空間を縮小することによって、計算時間の高速化を図る。図24の例では、点集合Aの要素数5の2倍の大きさ10を探索空間の大きさとして設定し処理を行っている。

#### 【0127】

図25にその点集合Bの分割アルゴリズムを示す。

順序付けられた点集合 $A = \{a_1, \dots, a_m\}$ 、 $B = \{b_1, \dots, b_i, \dots, b_j, \dots, b_n\}$ と点集合Bの部分集合 $B'$ において、以下の処理を行う。

#### 【0128】

処理1：点集合Aの要素数 $m$ を求める。

処理2：点集合 $B'$ の大きさを設定する関数 $f(x)$ により、 $B'$ の大きさ( $f(m)$ )を設定する。

#### 【0129】

処理3：点集合Bを分割する。

$$(a) j = i + f(m) - 1$$

$$(b) \text{点集合 } B' = \{b_i, b_{i+1}, \dots, b_{j-1}, b_j\}$$

処理4: 点  $a_1$  と  $b_j$  を対応付けた後, 点集合  $A, B'$  の残りの各要素の対応付けを行う。

【0130】

処理5:  $b_j$  が点集合  $B$  の最後の要素の時, 終了。

$b_j$  が点集合  $B$  の最後の要素ではない時,  $i = i + 1$  にして処理3に戻る。

【0131】

(2) 点集合  $A$  の空間的な大きさに応じた順序付けられた点集合  $B$  の分割1

図26に示すように, 点集合  $A$  の両端の2点の距離  $d$  を求め, この距離  $d$  により点集合  $B$  を分割し探索空間を縮小することによって, 計算時間の高速化を図る。特にこの手法は, 次の探索空間に移る際, 前の探索空間内で得られた解の位置を考慮して次の探索空間を設定することによって, 探索空間が重なることにより同じ解を何度も計算してしまうことを避けた手法である。

【0132】

図27は, 点集合  $A$  の空間的な大きさに応じた順序づけられた点集合  $B$  の分割アルゴリズムを示す図である。

順序付けられた点集合  $A = \{a_1, \dots, a_m\}$ ,  $B = \{b_1, \dots, b_i, \dots, b_j, \dots, b_n\}$  と点集合  $B$  の部分集合  $B'$  において, 以下の処理を行う。

【0133】

処理1: 点集合  $A, B$  の各点間の距離を計算し距離テーブル (図示省略) を作成する。

処理2: 点集合  $A$  の最初と最後の点 ( $a_1, a_m$ ) の距離を距離テーブルより求め,  $d$  とする。

【0134】

処理3: 点集合  $B$  を分割する。

(a)  $b_i$  ( $i = 1$  から) との距離が  $d \pm \alpha$  で,  $m \leq j - i \leq 2m$  を満たす  $b_j$  のうち  $j$  が最大のものを距離テーブルから求める。

## 【0135】

(b) 点集合  $B' = \{b_i, b_{i+1}, \dots, b_{j-1}, b_j\}$

処理4: 点集合A,  $B'$  の各要素の対応付けを行う。

処理5:  $b_j$  が点集合Bの最後の要素の時, 終了。

## 【0136】

処理6:  $b_j$  が点集合Bの最後の要素ではない時

①点集合A,  $B'$  の点に対応付けられた時,

$a_l$  と対応付けられた点を  $b_k$  とする時,  $i = k + 1$  として処理3に戻る。

## 【0137】

②点集合A,  $B'$  の点に対応付けられなかった時,

$i = i + 1$  として処理3に戻る。

(3) 点集合Aの空間的な大きさに応じた順序付けられた点集合Bの分割2

図28に示すアルゴリズムのように, 点集合Aの空間的な大きさに応じた順序付けられた点集合Bの分割を行うこともできる。この例でも点集合Aの両端の2点の距離を求め, この距離により点集合Bを分割し探索空間を縮小することによって, 計算時間の高速化を図っている。また, 次の探索空間に移る際, 探索のキーとなる点集合Aの要素数を考慮して次の探索空間を設定することによって, 探索空間が重なり同じ解を何度も計算してしまうことを避けている。

## 【0138】

順序付けられた点集合  $A = \{a_1, \dots, a_m\}$ ,  $B = \{b_1, \dots, b_i, \dots, b_j, \dots, b_n\}$  と点集合Bの部分集合  $B'$  において,

処理1: 点集合A, Bの各点間の距離を計算し距離テーブル(図示省略)を作成する。

## 【0139】

処理2: 点集合Aの最初と最後の点( $a_1, a_m$ )の距離を距離テーブルより求め,  $d$  とする。

処理3: 点集合Bを分割する。

## 【0140】

(a)  $b_i$  ( $i = 1$  から) との距離が  $d \pm a$  で、かつ  $m \leq j - i \leq 2m$  を満たす  $b_j$  のうち  $j$  が最大のものを距離テーブルから求める。

(b) 点集合  $B' = \{b_i, b_{i+1}, \dots, b_{j-1}, b_j\}$

処理 4 : 点集合  $A, B'$  の各要素の対応付けを行う。

【0 1 4 1】

処理 5 :  $b_j$  が点集合  $B$  の最後の要素の時、終了。

$b_i$  が点集合  $B$  の最後の要素ではない時、 $i = j - m + 1$  にして処理 3 に戻る。

【0 1 4 2】

以上により、立体構造を形成する点同士を対応付ける際、立体構造の探索空間を分割してから点同士の対応付けを行うため、短時間に点同士を対応付けることができるようになる。

【0 1 4 3】

本実施例による分子構造を重ね合わせて表示するシステムについても、最初に説明した実施例の場合と同様な構成で実現することができる。ただし、図 1 に示す組合せ生成処理手段 1 4 による対応付け方法が異なる。

【0 1 4 4】

具体例を説明するために、タンパク質 Trypsin のアミノ酸 (残基) 配列を図 2 9 (A) に、Elastase のアミノ酸配列を図 2 9 (B) に示す。図 2 9 は P D B に登録されているアミノ酸配列を抜粋したものである。図 2 9 に示したアミノ酸配列番号は、P D B に記載されているアミノ酸に対して単純に 1 から番号をふったものである。図 2 9 に示したアミノ酸番号とは異なっているが、以降の説明では図に示したアミノ酸番号を使用することにする。

【0 1 4 5】

図 2 9 に示した Trypsin, Elastase はセリンプロテアーゼと呼ばれるタンパク質分解酵素の仲間で、活性部位にヒスチジン、セリン、アスパラギン酸が不可欠な酵素である。これらの酵素の基質特異性は全く異なるが、構造および触媒機構などの点で類似していることから進化的には一群の酵素であろうと考えられている。



## 【0146】

図30(A)に、Trypsin のヒスチジン活性部位(36-41)をプローブとして、Elastaseのヒスチジン活性部位を検索した結果を示す。この結果によると、Trypsin の活性部位36-41に対して、Elastaseの41-46が対応付けられたことがわかる。

## 【0147】

また、図30(B)に、Trypsin のセリン活性部位(175-179)をプローブとして、Elastaseのセリン活性部位を検索した結果を示す。この結果によると、Trypsin の活性部位175-179に対して、Elastaseの186-190が対応付けられたことがわかる。これらの結果は、生化学的な実験で得られた結果と合致するものである。

## 【0148】

このように、本装置によってプローブとなる物質の立体構造を指定することによって、物質の立体構造を格納したデータベースからr.m.s.d.値が最小となる対応付けが算出できるため、対応付けられた部分を重ね合わせて表示することによって物質の最適な重ね合わせ表示を実現することができる。

## 【0149】

例えば、新薬の開発のように求める機能を持つ物質を開発する場合などには、物質の機能と構造の相関関係を解明するため、類似な立体構造を持つ物質を多数参照する必要が生じる。そのため、データベースから立体構造が類似した物質を簡単に取り出せる立体構造の検索システムが必要になるが、前述した例と同様に高速な検索システムを構成することができる。

## 【0150】

先に説明した実施例の手法により、組合せ生成処理手段14によって効率よく要素の組合せを生成し、類似性算出処理手段16によって類似する立体構造をデータベース21から抽出する。そして、結果出力処理手段18により、類似性算出処理手段16の結果に基づいて、類似な部分をアミノ酸配列名、およびアミノ酸番号で表し、類似性の尺度としてr.m.s.d.値を表示する。

## 【0151】

このように、本装置によってプローブ（検索キー）となる物質の立体構造を指定することによって、物質の立体構造を格納したデータベース 21 から r.m.s.d. 値が最小となる類似構造を検索することができる。

【0152】

【発明の効果】

以上説明したように、本発明によれば、立体構造を形成する点同士を自動的に重ね合わせることができるため、グラフィックシステムにより立体構造を自動的に重ね合わせて表示することや、立体構造が類似した構造をデータベースから自動的に検索することが可能になる。また、機能データベースを生成し、新しい物質の機能を予測することが可能になる。したがって、新薬の開発やタンパク質の研究等に寄与するところが大きい。

【図面の簡単な説明】

【図1】

本発明の原理説明図である。

【図2】

本発明に関係する r.m.s.d. 値の算出方法を示す図である。

【図3】

順序づけられていない点集合 A と B の対応付け説明図である。

【図4】

順序付けられていない点集合 A と B の対応付けアルゴリズムの例を示す図である。

【図5】

順序づけられた点集合 A と B の対応付け説明図である。

【図6】

順序付けられた点集合 A と B の対応付けアルゴリズムの例を示す図である。

【図7】

部分的に対応付けられた点集合 A と B の対応付け説明図である。

【図8】

幾何学的制約による候補の絞り込みを説明するための図である。

## 【図 9】

n i l 数を制限した点集合 A と B の対応付け説明図である。

## 【図 1 0】

タンパク質の例を示す図である。

## 【図 1 1】

本発明の実施例による検索結果の例を示す図である。

## 【図 1 2】

本発明の実施例による検索結果の例を示す図である。

## 【図 1 3】

立体構造検索結果の例を示す図である。

## 【図 1 4】

集合 A と B の部分集合の対応付け説明図である。

## 【図 1 5】

集合 A と B の部分集合の対応付けアルゴリズムの例を示す図である。

## 【図 1 6】

本発明の実施例に係る検索処理機構の説明図である。

## 【図 1 7】

タンパク質の二次構造のタイプの例とその定義説明図である。

## 【図 1 8】

二次構造の決定と部分集合への分割の処理の例を示す図である。

## 【図 1 9】

二次構造座標テーブルの構成例を示す図である。

## 【図 2 0】

図 1 6 に示す検索部での処理説明図である。

## 【図 2 1】

タンパク質 A のキーとした類似構造検索の出力例を示す図である。

## 【図 2 2】

図 2 1 に示す検索したタンパク質の立体構造を示す図である。

## 【図 2 3】

本実施例を適用する対象を説明するための直線の構造と非直線の構造の例を示す図である。

【図 2 4】

点集合 A の構成要素数に応じた順序づけられた点集合 B の分割の例を説明するための図である。

【図 2 5】

点集合 A の構成要素数に応じた順序づけられた点集合 B の分割アルゴリズムを示す図である。

【図 2 6】

点集合 A の空間的な大きさに応じた順序づけられた点集合 B の分割の例を説明するための図である。

【図 2 7】

点集合 A の空間的な大きさに応じた順序づけられた点集合 B の分割アルゴリズムを示す図である。

【図 2 8】

点集合 A の空間的な大きさに応じた順序づけられた点集合 B の他の分割アルゴリズムを示す図である。

【図 2 9】

アミノ酸配列の例を示す図である。

【図 3 0】

本実施例による検索結果の例を示す図である。

【符号の説明】

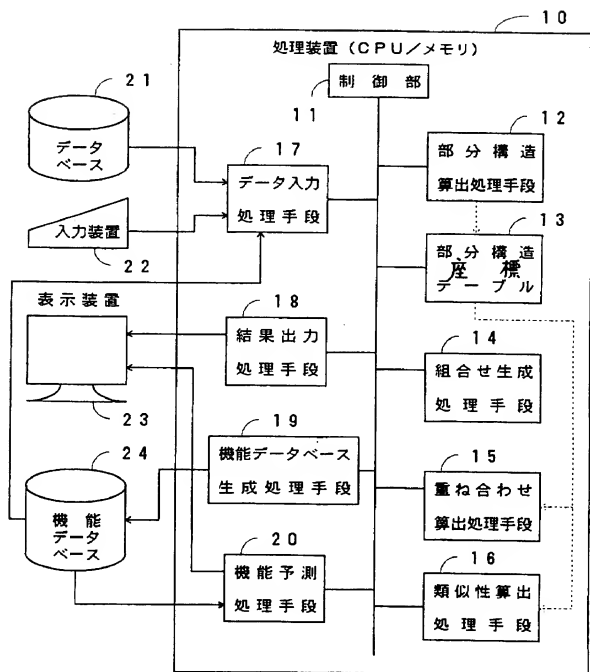
- 1 0 処理装置
- 1 1 制御部
- 1 2 部分構造算出処理手段
- 1 3 部分構造座標テーブル
- 1 4 組合せ生成処理手段
- 1 5 重ね合わせ算出処理手段
- 1 6 類似性算出処理手段

- 1 7 データ入力処理手段
- 1 8 結果出力処理手段
- 1 9 機能データベース生成処理手段
- 2 0 機能予測処理手段
- 2 1 データベース
- 2 2 入力装置
- 2 3 表示装置
- 2 4 機能データベース

【書類名】 図面

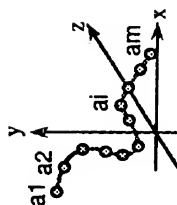
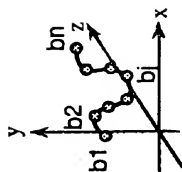
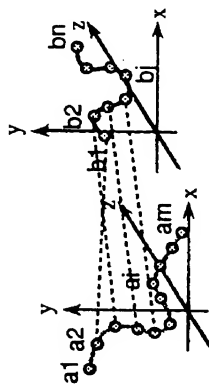
【図1】

本発明の原理説明図



【図2】

## r.m.s.d値の算出方法

(a) 点集合  $A = \{a_1, a_2, \dots, a_i, \dots, a_m\}$ (b) 点集合  $B = \{b_1, b_2, \dots, b_i, \dots, b_n\}$ 

(c) 点同士の対応付け

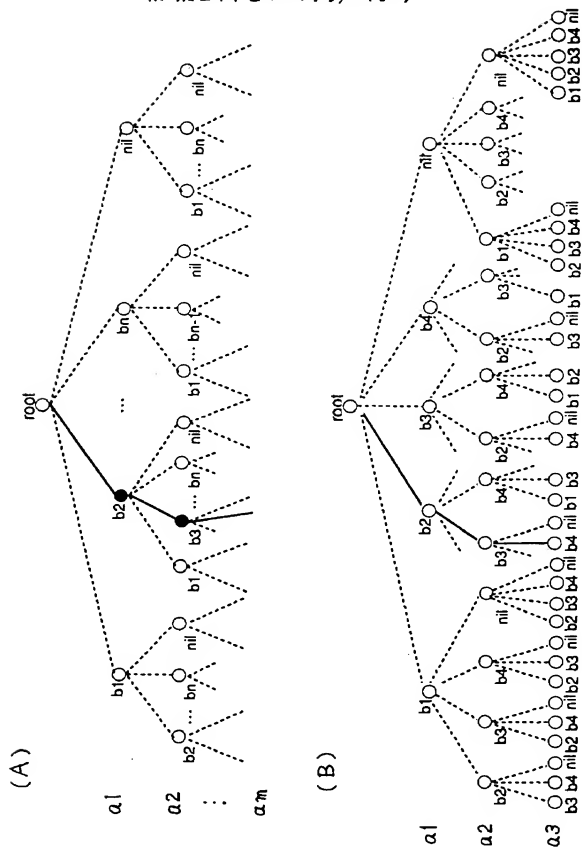
$$r.m.s.d. = \sqrt{\frac{\sum_{k=1}^n w_k (U_{b_k} - a_k)^2}{n}}$$



(d) r.m.s.d. 値

【図3】

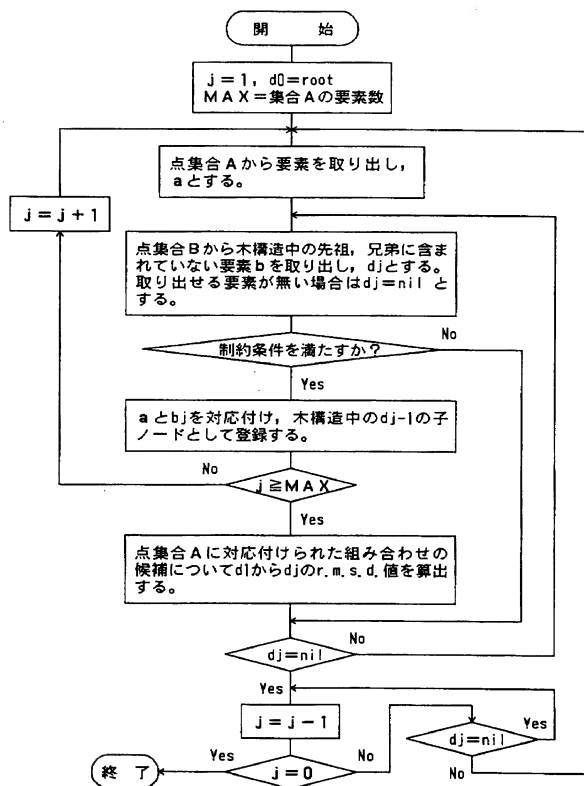
点集合AとBの対応付け





【図4】

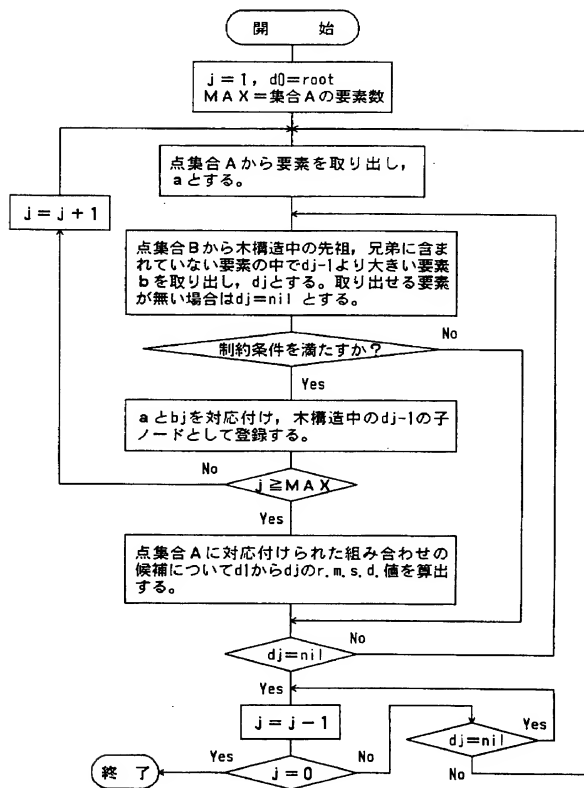
## 点集合AとBの対応付けアルゴリズム





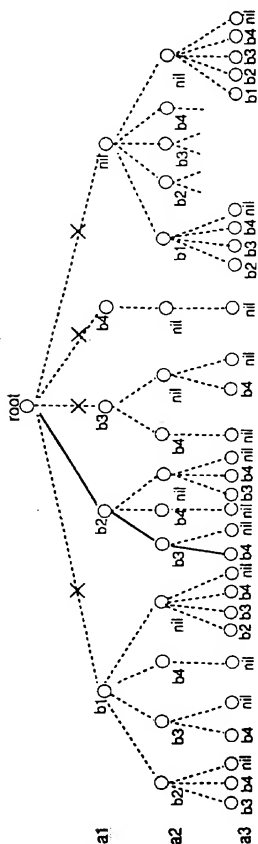
【図6】

順序付けられた点集合の対応付けアルゴリズム



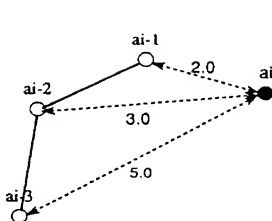
【図7】

部分的に対応付けられた点集合AとBの対応付け

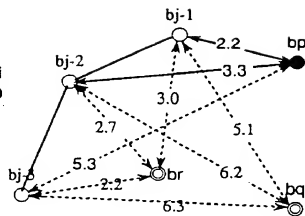


【図8】

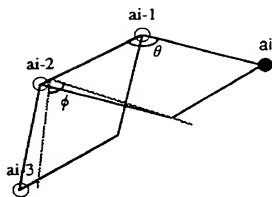
## 幾何学的制約



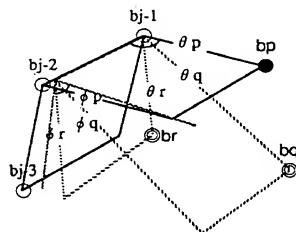
(A1) 点集合Aの距離関係



(A2) 点集合Bの距離関係



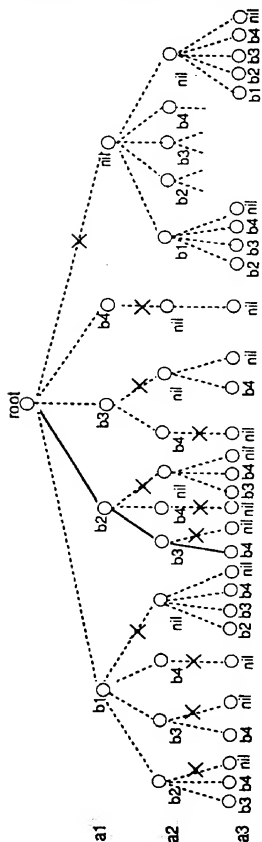
(B1) 点集合Aの角度関係



(B2) 点集合Bの角度関係

【図 9】

nil 数を制限した点集合 A と B の対応付け



【図10】

## タンパク質の例

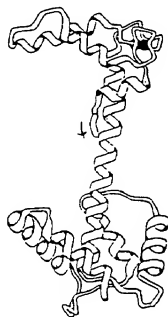
1	T E E Q I A E F K E	A F S L F D K D G D
21	G T I T T K E L G T	V M R S L G Q N P T
41	E A E L Q D M I N E	V D A D G N G T I D
61	F P E F L T M M A R	K M K D T D S E E E
81	I R E A F R V F D K	D G N G Y I S A A E
101	L R H V M T N L G E	K L T D E E V D E M
121	I R E A N I D G D G	Q V N Y E E F V Q M
141	M T A	

(A) Calmodulinのアミノ酸残基配列 (PDBより抜粋)

1	A M D Q Q A E A R A	F L S E E M I A E F
21	K A A F D M F D A D	G G G D I S T K E L
41	G T V M R M L G Q N	P T K E E L D A I I
61	E E V D E D G S G T	I D F E E F L V M M
81	V R Q M K E D A K G	K S E E E L A D C F
101	R I F D K N A D G F	I D I E E L G E I L
121	R A T G E H V T E E	D I E D L M K D S D
141	K N N D G R I D F D	E F L K M M E G V Q
161		

(B) Troponin Cのアミノ酸残基配列 (PDBより抜粋)

(C) トロポニンC



【図11】

## 検索結果の例1 (単一サイトの検索)

Probe site = 81-108 in Calmodulin

96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	
L	A	D	C	F	R	I	F	D	K	N	A	D	G	F	< target >
I	R	E	A	F	R	V	F	D	K	D	G	N	G	Y	< probe >
111	112	113	114	115	116	117	118	119	120	121	122	123			
I	D	I	E	E	L	G	E	I	L	R	A	T			< target >
I	S	A	A	E	L	R	H	V	M	T	N	L			< probe >

0.567034

probe= 28, target= 160, combination= 2 Time= 1sec Best rmsd= 0.567034



【図12】

## 検索結果の例2 (複数サイトの検索)

Probe site = 81-108 and 117-143 in Calmodulin

```

96 97 98 99 100 101 102 103 104 105 106 107 108 109 110
L A D C F R I F D K N A D G F < target >
I R E A F R V F D K D G N G Y < probe >
111 112 113 114 115 116 117 118 119 120 121 122 123
I D I E E L G E I L R A T < target >
I S A A E L R H V M T N L < probe >
132 133 134 135 136 137 138 139 140 141 142 143 144 145
I E D L M K D S D K N N D G < target >
V D E M I R E A N I D G D G < probe >
146 147 148 149 150 151 152 153 154 155 156 157 158
R I D F D E F L K M M E G < target >
Q V N Y E E F V Q M M T A < probe >
0.823665

```

probe= 55, target= 160, combination= 1 Time= 2sec Best rmsd= 0.823665

【図13】

## 立体構造検索結果の例

```

===== ATP/GTP binding site =====
Probe= /usr/local/grease/db/pdb/distr/tape3/pdb1efm.ent (elongation factor)

7 8 9 10 11 12 13 14
G H V D H G K T < probe >

-----
8 9 10 11 12 13 14 15
G A P G S G K G < target >
G H V D H G K T < probe >
0.648732

/usr/local/grease/db/pdb/distr/tape1/pdb1ak3.ent:unit-A (adenylate kinase)
probe= 8, target= 214, combination= 1 Best rmsd= 0.648732

:
:
:
10 11 12 13 14 15 16 17
G A G G V G K S < target >
G H V D H G K T < probe >
0.421770

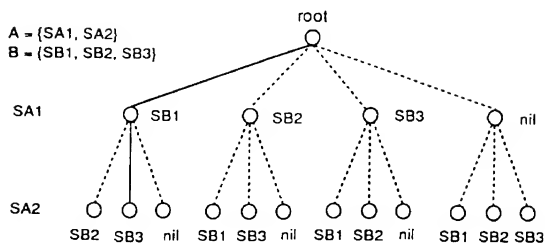
/usr/local/grease/db/pdb/distr/tape6/pdb2p21.ent (ras protein)
probe= 8, target= 166, combination= 1 Best rmsd= 0.421770

Total/Search/Match= 824/744/9, Total time= 3550sec
-----

```

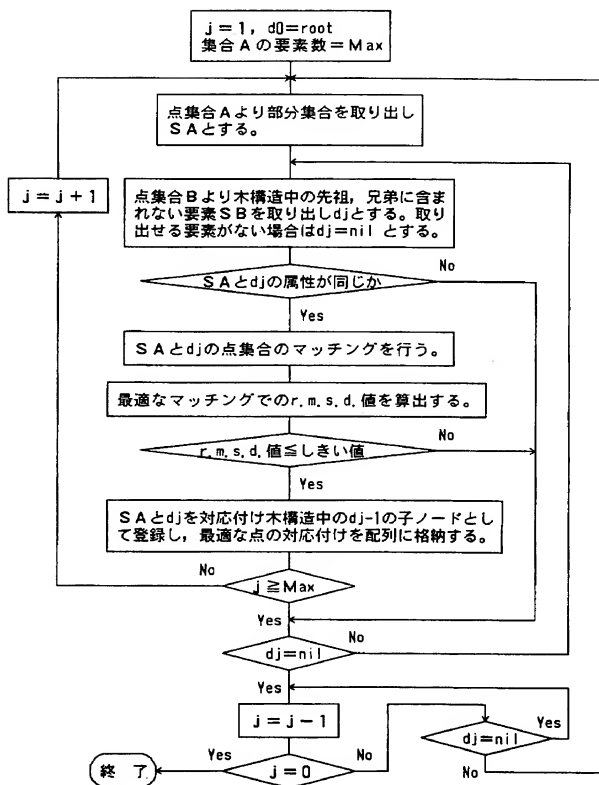
【図14】

集合AとBの部分集合の対応付け



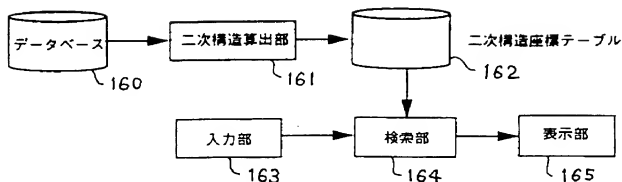
【図15】

点集合AとBの部分集合の対応付けアルゴリズム



【図16】

## 検索処理機構



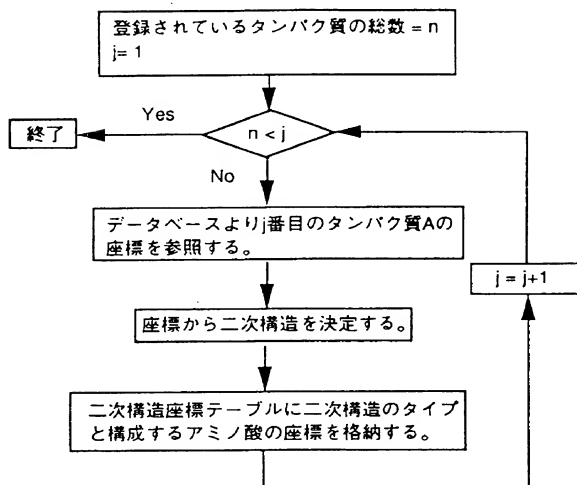
【図17】

## 二次構造のタイプの例とその定義

タイプ	定義
310ヘリックス	i番目の残基のカルボニル基とi+3番目の残基のアミド基が水素結合しながら達なった構造
$\alpha$ ヘリックス	i番目の残基のカルボニル基とi+4番目の残基のアミド基が水素結合しながら達なった構造
平行 $\beta$ シート	i-1番目の残基のカルボニル基とj番目の残基のアミド基、かつj番目の残基のカルボニル基とi+1番目の残基のアミド基、またはj-1番目の残基のカルボニル基とi番目の残基のアミド基、かつi番目の残基のカルボニル基とj+1番目の残基のアミド基が水素結合している構造
3ターン	i番目の残基のカルボニル基とi+2番目の残基のアミド基が水素結合している構造
⋮	⋮

【図18】

## 二次構造の決定と部分集合への分割



【図19】

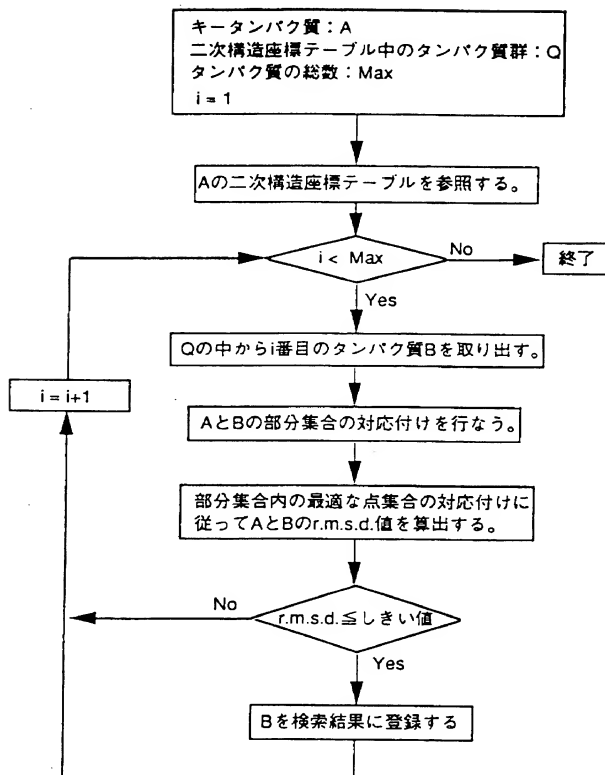
162

二次構造座標テーブル

SubSet	座標	タイプ
S1	{ X1, X2, X3, X4, ..... Xa }	$\alpha$ ヘリックス
S2	{ Xa+1, Xa+2, ..... Xb }	$\alpha$ ヘリックス
S3	{ Xb+1, Xb+2, ..... Xc }	$\beta$ シート
S4	{ Xc+1, Xc+2, ..... Xd }	$\beta$ シート
⋮	⋮	⋮
Sn	{ Xl+1, Xl+2, ..... Xm }	3ターン

【図20】

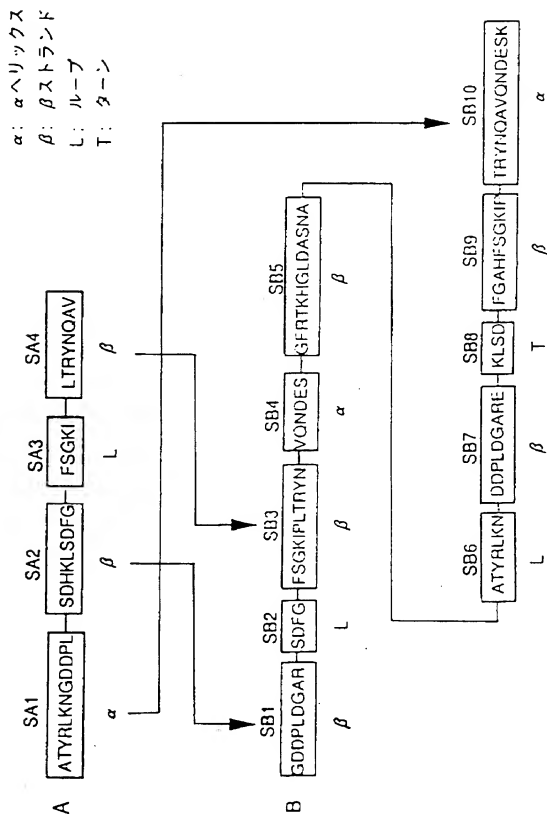
## 検索部での処理





【図21】

タンパク質Aをキーとした類似構造検索の出力例



【図22】

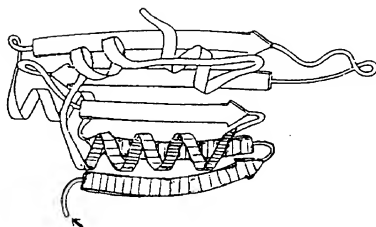
検索したタンパク質の立体構造

(A)



キータンパク質 : A

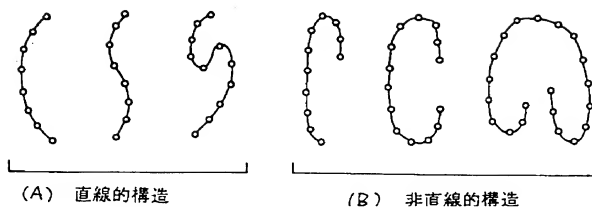
(B)



類似構造タンパク質 : B

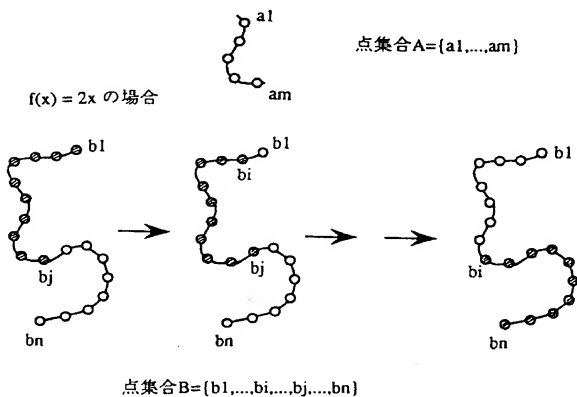
【図23】

## 直線の構造と非直線の構造



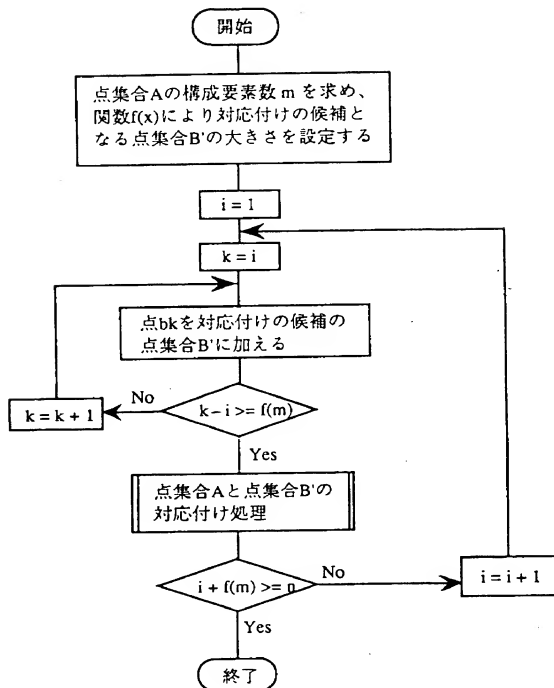
【図24】

点集合Aの構成要素数に応じた順序づけられた点集合Bの分割



【図 2 5】

## 点集合Bの分割アルゴリズム

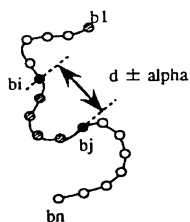


【図26】

点集合Aの空間的な大きさに応じた順序づけられた点集合Bの分割



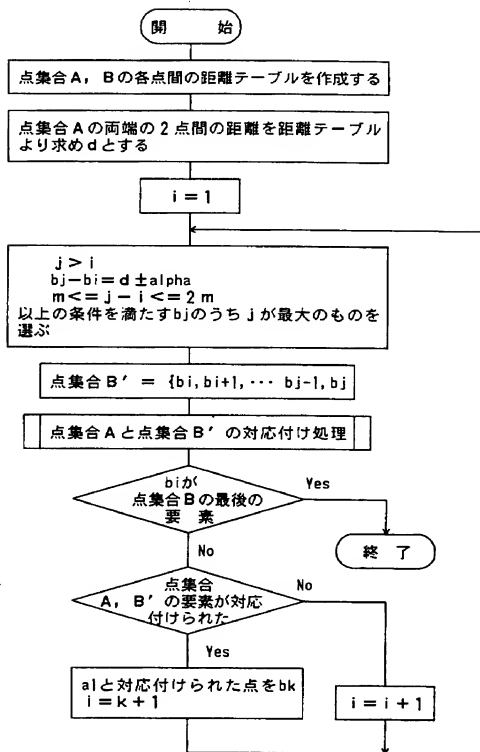
点集合A={a1,a2,...,am}



点集合B={b1,...,bi,...,bj,...,bn}

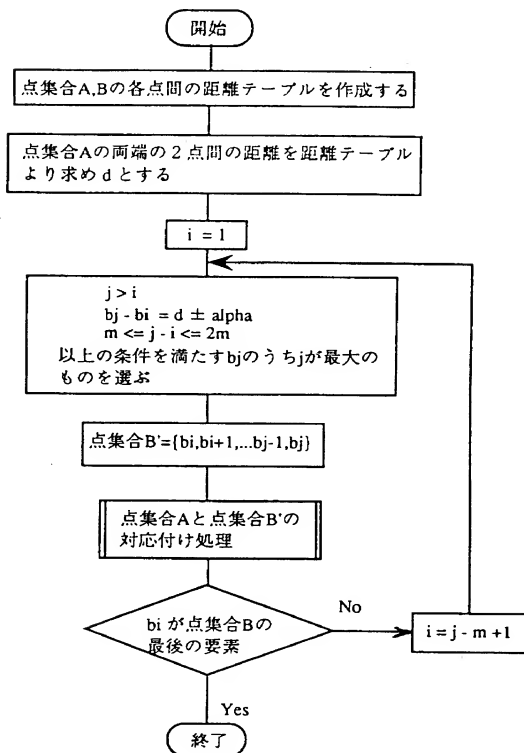
【図27】

## 点集合Bの分割アルゴリズム(その1)



【図28】

## 点集合Bの分割アルゴリズム (その2)



【図29】

## アミノ酸配列の例

1	I	V	G	G	Y	T	C	G	A	N	T	V	P	Y	Q	V	S	L	N	S
21	G	Y	H	F	C	G	G	S	L	I	N	S	Q	W	V	V	S	A	A	H
41	C	Y	K	S	G	I	Q	V	R	L	G	E	D	N	I	N	V	V	E	G
61	N	E	Q	F	I	S	A	S	K	S	I	V	H	P	S	Y	N	S	N	T
81	L	N	N	D	I	M	L	I	K	L	K	S	A	A	S	L	N	S	R	V
101	A	S	I	S	L	P	T	S	C	A	S	A	G	T	Q	C	L	I	S	G
121	W	G	N	T	K	S	S	G	T	S	Y	P	D	V	L	K	C	L	K	A
141	P	I	L	S	D	S	S	C	K	S	A	Y	P	G	Q	I	T	S	N	M
161	F	C	A	G	Y	L	E	G	G	K	D	S	C	Q	G	D	S	G	G	P
181	V	V	C	S	G	K	L	Q	G	I	V	S	W	G	S	G	C	A	Q	K
201	N	K	P	G	V	Y	T	K	V	C	N	Y	V	S	W	I	K	Q	T	I
221	A	S	N																	

(A) Trypsinのアミノ酸配列 (PDBより抜粋)

1	V	V	G	G	T	E	A	Q	R	N	S	W	P	S	Q	I	S	L	Q	Y
21	R	S	G	S	S	W	A	H	T	C	G	G	T	L	I	R	Q	N	W	V
41	M	T	A	A	H	C	V	D	R	E	L	T	F	R	V	V	V	G	E	H
61	N	L	N	Q	N	N	G	T	E	Q	Y	V	G	V	Q	K	I	V	V	H
81	P	Y	W	N	T	D	D	V	A	A	G	Y	D	I	A	L	L	R	L	A
101	Q	S	V	T	L	N	S	Y	V	Q	L	G	V	L	P	R	A	G	T	I
121	L	A	N	N	S	P	C	Y	I	T	G	W	G	L	T	R	T	N	G	Q
141	L	A	Q	T	L	Q	Q	A	Y	L	P	T	V	D	Y	A	I	C	S	S
161	S	S	Y	W	G	S	T	V	K	N	S	M	V	C	A	G	G	D	G	V
181	R	S	G	C	Q	G	D	S	G	G	P	L	H	C	L	V	N	G	Q	Y
201	A	V	H	G	V	T	S	F	V	S	R	L	G	C	N	V	T	R	K	P
221	T	V	F	T	R	V	S	A	Y	I	S	W	I	N	N	V	I	A	S	N

(B) Elastaseのアミノ酸配列 (PDBより抜粋)



【図30】

## 検索結果の例

Key site number 36 - 41 in Trypsin

41 42 43 44 45 46

M T A A H C &lt; target &gt;

V S A A H C &lt; probe &gt;

d = 12.070038 [A]

r.m.s.d. = 0.061077 [A]

The number of atoms in a probe = 6

The number of atoms in PDB = 240

The number of combination = 1

Time = 1sec

(A) Histidine active site の検索結果

Key site number 175 - 179 in Trypsin

186 187 188 189 190

G D S G G &lt; target &gt;

G D S G G &lt; probe &gt;

d = 8.922721 [A]

r.m.s.d. = 0.092879 [A]

The number of atoms in a probe = 5

The number of atoms in PDB = 240

The number of combination = 1

Time = 1sec

(B) Serine active site の検索結果

【書類名】 要約書

【要約】

【目的】物質の二つの立体構造を重ね合わせたり、類似部分を抽出したりする立体構造処理装置に関し、重ね合わせ表示や、データベースからの類似構造の検索を自動化し、化学物質やタンパク質の構造解析における構造比較、類似構造検索等の作業を効率化する手段を提供することを目的とする。

【構成】立体構造を形成する二つの点集合の各要素を対応付ける際に、組合せ生成処理手段14により、所定の制約条件を満たす要素の組合せを生成し、重ね合わせ算出処理手段15によってその組合せの中で各点間の平均二乗距離の平均値が小さくなる重ね合わせ方法を算出する。また、検索の際には、類似性算出処理手段16により、各点間の平均二乗距離の平均値が小さくなる類似構造をデータベース21から自動的に探し出す。

【選択図】 図 1



【書類名】	職権訂正データ
【訂正書類】	特許願

## &lt; 認定情報・付加情報 &gt;

## 【特許出願人】

【識別番号】	000005223
【住所又は居所】	神奈川県川崎市中原区上小田中 1 0 1 5 番地
【氏名又は名称】	富士通株式会社

## 【代理人】

【識別番号】	100087848
【住所又は居所】	東京都荒川区西日暮里 4 丁目 1 7 番 1 号 佐原マン ション 3 F C 小笠原特許事務所
【氏名又は名称】	小笠原 吉義

特願平 0 4 - 3 3 1 7 0 3

出 願 人 履 歴 情 報

識別番号

[ 0 0 0 0 0 5 2 2 3 ]

- |          |                             |
|----------|-----------------------------|
| 1. 変更年月日 | 1 9 9 0 年   8 月 2 4 日       |
| [変更理由]   | 新規登録                        |
| 住 所      | 神奈川県川崎市中原区上小田中 1 0 1 5 番地   |
| 氏 名      | 富士通株式会社                     |
| 2. 変更年月日 | 1 9 9 6 年   3 月 2 6 日       |
| [変更理由]   | 住所変更                        |
| 住 所      | 神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号 |
| 氏 名      | 富士通株式会社                     |